Nonlinear Time Series Analysis and its Application

非线性时间序列分析及其应用

王海燕　卢　山　著



科学出版社
www.sciencep.com

(O-2593.0101)

非线性时间序列分析及其应用

Nonlinear Time Series Analysis and its Application



ISBN 7-03-018035-6
定 价：30.00 元

东南大学科技出版基金资助

非线性时间序列分析及其应用

王海燕　卢　山　著

科　学　出　版　社

北　京

内 容 简 介

本书以来自于确定性非线性系统的观测或实验时间序列为研究对象，在对问题的背景和意义进行分析的基础上，根据目前国内外关于单变量非线性时间序列分析的相关文献，总结了单变量非线性时间分析的基本流程，对单变量非线性时间序列分析的基本方法进行了详细综述。由于实际问题中常常可以获得多变量时间序列，本书把单变量非线性时间序列分析方法推广到多变量非线性时间序列的情形，着重研究了基于多变量时间序列的系统非线性性检验方法、多变量时间序列相空间重构方法和多变量非线性时间序列的预测方法等，最后把这些方法应用到证券市场的指数时间序列中。

本书自成体系，可作为系统工程、管理科学、金融工程、应用数学、生物医学工程、信号处理等专业高年级本科生、研究生和从事相关领域研究的科技工作者的参考书。

图书在版编目(CIP)数据

非线性时间序列分析及其应用／王海燕，卢山著．—北京：科学出版社，2006

ISBN 7-03-018035-6

Ⅰ．非…　Ⅱ．①王…②卢…　Ⅲ．非线性-时间序列分析　Ⅳ．O211.61

中国版本图书馆 CIP 数据核字(2006)第 108[illegible]09 号

责任编辑：何舒民／责任校对：赵　燕

责任印制：吕春珉／封面设计：耕者设计工作室

科学出版社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

双青印刷厂 印刷

科学出版社发行　各地新华书店经销

*

2006 年 11 月第 一 版　　开本：B5（720×1000）

2006 年 11 月第一次印刷　　印张：11 1/2

印数：1—2 000　　字数：230 000

定价：30.00 元

（如有印装质量问题，我社负责调换<新欣>）

销售部电话 010-62136131　编辑部电话 010-62137026(BA03)

前　言

近年来，非线性科学得到了快速发展。实际问题中，面临大量复杂而无法直接建立解析数学模型的非线性系统，对这些非线性系统可以通过实验或观测的手段获得非线性时间序列，这些非线性时间序列中蕴含着丰富的系统动力学信息，如何提取这些信息并应用到实际问题中去解释、分析，甚至控制原来的复杂系统是非线性时间序列应用的一个重要方面。从 20 世纪 80 年代以来，非线性时间序列分析的相关理论已取得了许多重大成果，在实际复杂系统中也得到了有效的应用，为解决实际问题中像经济系统、生理系统、电力系统等这些用传统的确定性和随机性方法都无法很好刻画的复杂系统提供了一种全新的研究工具。

基于混沌理论的非线性时间序列分析目前在许多领域都引起广泛兴趣，其中的某些成果已在国内外得到应用，但目前国内系统介绍这方面的著作还很少，应用工作多是某一具体方法在实际问题中的应用，缺乏对理论和应用的系统总结。作者在给相关专业的博士生和硕士生上课时对非线性时间序列分析方法进行过介绍，引起了学生的广泛兴趣，所以在本书第二章对目前得到广泛应用的单变量非线性时间序列的基本流程进行了总结和归纳，对其分析方法进行了详细综述。同时作者结合近年来把单变量非线性时间序列分析方法推广到多变量非线性时间序列情形的研究实践，在本书第三至第五章详细介绍了基于多变量时间序列的系统非线性性检验方法、多变量时间序列相空间重构方法和多变量非线性时间序列的预测方法等，并在第六章把这些方法应用到证券市场的指数时间序列中，且进行了实证分析。本书介绍的非线性时间序列分析方法，也可应用到其他领域，例如，经济系统的汇率时间序列、生理系统的心(或脑)电图时间序列、电力系统电力负荷时间序列、水文预报的水沙(或流量)时间序列、机械系统故障诊断的振动信号检测时间序列等。

本书内容自成体系，学术上力求完整严谨，应用上力求通俗易懂，把复杂问题用简单的语言说清楚，使对这一领域感兴趣的读者通过对本书的阅读能很快进入这一领域并掌握非线性时间序列分析这一工具。对理论性较强的内容和相关的应用成果，本书标注了大量的参考文献，使不同学科领域的读者可以通过本书对非线性时间序列分析这一领域有一个较全面的了解。

本书在写作过程中参考或引用了许多学者的资料，作者已尽可能在文中作了标注，并在参考文献中列出，在此，谨对他们表示衷心的感谢。若某些引用的资料由于作者疏忽没有标注其出处，作者深表歉意。

本书由王海燕负责全书结构的策划和具体的撰写，卢山负责相关仿真分析的具体计算，研究生汤龙坤、朱梅、钱志强、赵巍、方芬等在仿真分析中也做了大量具体工作，借此机会向他们表示衷心的感谢。

本书得到了东南大学科技出版基金的资助，在此表示衷心的感谢。

由于非线性时间序列分析及其应用是理论和实践都很强的研究领域，尚有许多问题有待深入分析和研究，因此书中难免有不足之处，恳请读者批评指正。

王海燕

2006 年 6 月于东南大学

目　　录

第一章 绪　论

对实际问题中动力系统进行研究的一种常用方法是通过对系统本身的性质分析,用差分方程或微分方程建立系统的数学模型,然后通过对数学模型的分析来研究系统的演化规律。从数学上看,这是一个正问题,系统的演化规律能否正确刻画取决于对系统所做的假设和在此假设基础上所建立的数学模型的正确性。实际问题中有些系统很难通过这种方法来建立数学模型,如像经济系统、生理系统、社会系统等要建立精确的解析模型是非常困难的。但对这些系统通过实验或观测可以获得包含系统演化信息的时间序列,需要研究的是能否通过这些时间序列本身去获取所研究系统的相关信息以及如何通过这些时间序列去揭示所研究系统的演化行为。从数学上看,这是一个反问题,它对所研究系统本身的背景信息和知识要求较少,只是从获取的时间序列本身出发进行研究,因此研究起来会更困难,但对那些无法建立精确解析模型的实际系统提供了一种新的研究方法。时间序列分析法正是从这样的角度去研究复杂系统演化行为的一种方法。

1.1 时间序列的含义及分类

从纵向看,时间序列是指存在于自然科学或社会科学中的某一变量或指标的数值或观测值,按照其出现时间的先后次序,以相同的或不同的间隔时间排列的一组数值。它是某一现象或若干现象在不同时刻上的状态所形成的数据,反映的是现象以及现象之间关系的发展变化规律。从横向看,时间序列也可以是若干相关现象在某一时间点上所处的状态按一定顺序排序的一组数据,反映的是一定时间、地点条件下各相关现象之间存在的内在数值联系。因此,从系统的意义上看,时间序列就是某一系统在不同时间(地点、条件等)的响应。

时间序列是指所研究系统的历史行为的客观记录,因而它包含了系统结构特征及其运行规律,所以可通过对时间序列的研究来认识所研究系统的结构特征,揭示所研究系统的运行规律,进而用以预测、控制其未来行为,修正和重新设计系统使之按照希望的结构运行。

时间序列根据研究的出发点不同,可以有不同的分类。

1) 按研究系统复杂程度不同,可以分为线性时间序列和非线性时间序列。线性时间序列是指从线性系统通过观察或实验获取的时间序列。而非线性时间序列是指从非线性系统通过观察或实验获取的时间序列。线性时间序列比较简单,可以

通过线性回归等方法建立数学模型进行分析；非线性时间序列比较复杂，直接建立数学模型比较困难，有时甚至具有混沌特征或分形特征。

2）按研究系统的确定性程度不同，可以分为随机时间序列和确定性时间序列。随机时间序列是指从随机系统通过观察或实验获取的时间序列，一般用统计分析方法进行研究。确定性时间序列是指从确定性系统通过观察或实验获取的时间序列，可以通过建立确定性数学模型进行研究，但当确定性系统具有混沌特征时，对这样的时间序列进行分析会非常复杂。

3）按研究系统的观察变量多少不同，可以分为单变量时间序列和多变量时间序列。单变量时间序列是指从系统中只观察一个量所得到的时间序列。多变量时间序列是指从系统中同时观察多个变量所得到的多维时间序列。多变量时间序列中观察量之间可能是相关的，这样的多变量时间序列就有冗余。

本书研究的时间序列主要是指来自于非线性确定性系统的实测单变量或多变量时间序列。

1.2 非线性时间序列的例子

来自于非线性确定性系统的时间序列具有各种各样的特性，有些随时间的演化会趋向于一个稳定的状态，有些随时间的演化会周期性变化，还有些随时间的演化会表现出看起来是随机的行为。

1. *趋向于稳定状态的时间序列*

例 1.1 考虑以下 Logistic 模型

$$x_{n+1} = Rx_n(1 - x_n) \tag{1.1}$$

当 $R=1.5$ 时，不管初始状态 x_0 在何处，随时间的演化，系统都将单调地趋向于 $\frac{1}{3}$，如图 1.1 所示（$x_0=0.7$）。

当 $R=2.9$ 时，不管初始状态 x_0 在何处，随时间的演化，系统都将交替地趋向于 $\frac{19}{29}$，如图 1.2 所示（$x_0=0.7$）。

例 1.2 弹簧振子在策动力、弹性力和阻尼力的作用下做受迫振动，其振动方程为

$$\frac{d^2x}{dt^2} + 2\beta\frac{dx}{dt} + \omega_0^2 x = f_0\cos\alpha t \tag{1.2}$$

取物体所受的最大策动力 $f_0=100$，策动力的频率 $\alpha=1$，振动物体的固有频率 $\omega_0=0$，阻尼因子 $\beta=0.1$。当初值 $x(0)=5$，$x'(0)=0$ 时，随时间的演化，系统将趋于稳定状态，如图 1.3 所示。

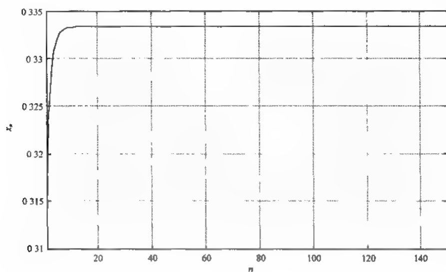


图 1.1 Logistic 模型中 $R=1.5$ 时的时间历程图

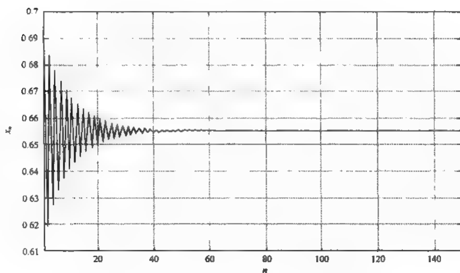


图 1.2 Logistic 模型中 $R=2.9$ 时的时间历程图

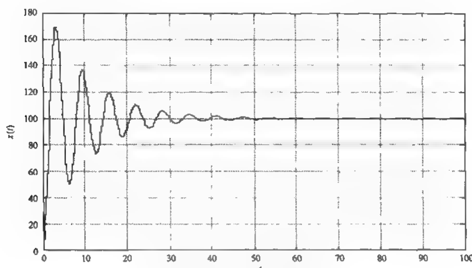


图 1.3　具有阻尼的振动方程的时间历程图

2. 趋向于周期状态的时间序列

例 1.3　考虑以下 Logistic 模型

$$x_{n+1} = 3.3x_n(1 - x_n) \tag{1.3}$$

不管初始状态 x_0 在何处，随时间的演化，系统都将在 0.48 和 0.82 两个状态之间周期性地变化，如图 1.4 所示（$x_0=0.7$）。

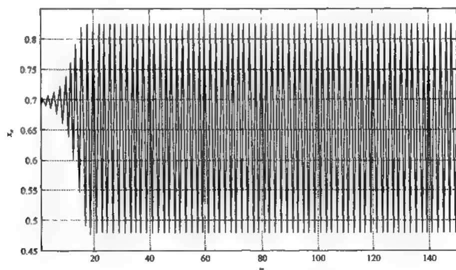


图 1.4　Logistic 模型中 $R=3.3$ 时的时间历程图

例 1.4 弹簧振子在策动力、弹性力和阻尼力的作用下做受迫振动，其振动方程为

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2}+2\beta\frac{\mathrm{d}x}{\mathrm{d}t}+\omega_0^2x=f_0\cos\alpha t \tag{1.4}$$

取物体所受的最大策动力 $f_0=100$，策动力的频率 $\alpha=1$，振动物体的固有频率 $\omega_0=1$，阻尼因子 $\beta=0.1$。当初值 $x(0)=5, x'(0)=0$ 时，随时间的演化，系统将趋于周期性变化状态，如图 1.5 所示。

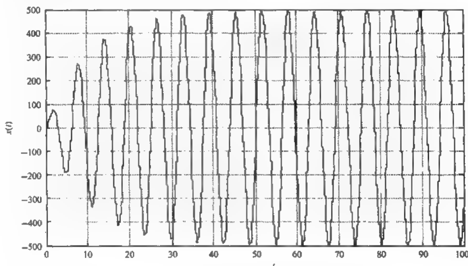


图 1.5 无阻尼的振动方程的时间历程图

3. 混沌时间序列

例 1.5 考虑以下 Logistic 模型

$$x_{n+1}=4x_n(1-x_n) \tag{1.5}$$

随时间的演化，系统将出现不规则的振荡，看起来好像是随机的，如图 1.6 所示。

在 Logistic 模型(1.5)中，初值的微小变化随着时间的演化会不断放大，表 1.1是 Logistic 模型(1.5)的三个非常接近的初值演化 50 次后的结果。可以看到刚开始时演化结果非常接近，但随着演化次数的增加，演化结果不断分离，最后甚至变得毫无关系，表明系统对初值具有非常敏感的依赖性，也说明这样的系统只能进行短期预测，要进行较长时间的预测会变得不正确。

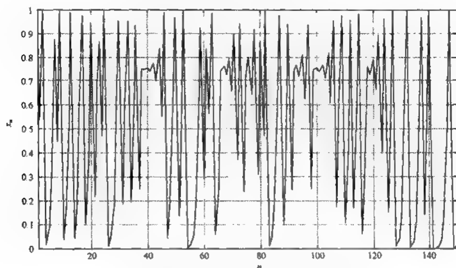


图 1.6 Logistic 模型中 $R=4$ 时的时间历程图

表 1.1 Logistic 模型(1.5)在不同初值下的演化结果

n	$x_n^{(1)}$	$x_n^{(2)}$	$x_n^{(3)}$
1	0.100 000 000 000 000 0	0.100 000 010 000 000 0	0.100 000 020 000 000 0
2	0.360 000 000 000 000 0	0.360 000 031 999 999 6	0.360 000 063 999 998 4
3	0.921 600 000 000 000 0	0.921 600 035 839 995 4	0.921 600 071 679 981 8
4	0.289 013 760 000 000 1	0.289 013 639 118 858 4	0.289 013 518 237 736 8
5	0.821 939 226 122 649 8	0.821 939 022 088 530 7	0.821 938 818 054 328 7
6	0.585 405 387 334 197 4	0.585 421 064 226 722 4	0.585 421 589 719 127 5
7	0.970 813 326 249 438 0	0.970 812 967 145 496 7	0.970 812 608 039 428 2
8	0.113 339 247 303 761 2	0.113 340 599 870 613 8	0.113 341 952 444 447 2
9	0.401 973 849 297 512 3	0.401 978 033 166 332 8	0.401 982 217 042 111 4
10	0.961 563 495 113 812 8	0.961 566 776 072 237 8	0.961 570 056 896 080 9
11	0.147 836 559 913 285 3	0.147 824 444 905 122 6	0.147 812 330 307 794 6
12	0.503 923 645 865 163 6	0.503 889 513 574 460 0	0.503 855 381 267 096 2
13	0.999 938 420 012 499 1	0.999 939 486 736 616 4	0.999 940 544 141 141 3
14	0.000 246 304 781 624 1	0.000 242 038 406 114 3	0.000 237 809 295 438 2
15	0.000 984 976 462 314 7	0.000 967 919 294 097 3	0.000 951 010 968 708 6
16	0.003 936 025 134 733 6	0.003 867 929 705 349 5	0.003 800 426 187 384 2
17	0.015 682 131 363 489 3	0.015 411 875 300 575 8	0.015 143 931 792 713 7
18	0.061 744 808 477 550 3	0.060 697 397 601 181 4	0.059 658 372 490 285 4

续表

n	$x_n^{(1)}$	$x_n^{(2)}$	$x_n^{(3)}$
19	0.231 729 548 414 483 7	0.228 052 894 102 501 9	0.224 397 004 328 382 9
20	0.712 123 859 224 412 5	0.704 179 086 375 899 8	0.696 171 955 107 322 4
21	0.820 013 873 390 966 5	0.833 243 602 746 811 4	0.846 066 256 117 482 7
22	0.590 364 483 349 241 7	0.555 794 804 913 301 4	0.520 952 585 507 315 6
23	0.967 337 040 596 098 5	0.987 547 758 978 746 6	0.998 243 956 642 234 5
24	0.126 384 361 947 522 4	0.049 188 730 859 208 1	0.007 011 838 677 964 5
25	0.441 645 420 010 560 2	0.187 076 798 462 674 1	0.027 850 691 185 274 7
26	0.986 378 971 977 024 3	0.608 316 279 758 520 4	0.108 300 120 743 108 5
27	0.053 741 982 474 292 1	0.953 070 334 157 095 7	0.386 284 818 360 546 6
28	0.203 415 127 176 099 9	0.178 909 089 227 110 4	0.948 275 429 858 824 5
29	0.648 149 652 848 124 1	0.587 602 508 076 145 0	0.196 196 555 939 544 4
30	0.912 206 721 443 921 2	0.969 303 202 315 075 8	0.630 813 869 508 022 5
31	0.320 342 475 185 814 1	0.119 018 017 187 260 0	0.931 550 926 177 352 3
32	0.870 892 695 110 560 7	0.419 410 915 088 292 4	0.255 055 192 461 877 9
33	0.449 754 434 854 498 7	0.974 021 597 572 374 4	0.760 008 165 040 449 3
34	0.989 901 532 732 836 8	0.101 214 100 139 735 8	0.729 583 016 449 193 9
35	0.039 985 952 904 069 1	0.363 879 224 290 557 4	0.789 166 554 232 356 7
36	0.153 548 305 897 690 6	0.925 884 537 681 038 5	0.665 530 815 653 542 0
37	0.519 884 894 614 559 3	0.274 489 442 256 832 4	0.890 398 196 276 292 4
38	0.998 418 363 864 671 5	0.796 579 953 385 461 8	0.390 356 993 376 870 0
39	0.006 316 538 249 855 6	0.648 161 324 999 509 2	0.951 913 644 394 561 1
40	0.025 106 558 377 574 7	0.912 192 887 097 559 3	0.183 096 232 040 104 9
41	0.097 904 876 416 032 6	0.320 388 095 304 714 9	0.598 288 007 411 283 7
42	0.353 278 046 359 975 8	0.870 958 254 766 927 3	0.961 357 870 396 477 7
43	0.913 890 673 280 218 5	0.449 559 892 881 101 7	0.148 595 661 692 907 6
44	0.314 778 042 286 589 9	0.989 823 182 375 336 2	0.506 059 964 075 818 1
45	0.862 771 305 523 246 9	0.040 293 000 030 792 6	0.999 853 107 341 599 2
46	0.473 587 919 555 836 3	0.154 677 896 717 244 7	0.000 587 484 323 791 0
47	0.997 209 608 026 444 0	0.523 010 579 937 496 3	0.002 348 556 743 841 2
48	0.011 130 422 744 759 7	0.997 882 052 843 760 4	0.009 372 164 100 248 7
49	0.044 026 145 737 130 6	0.008 453 845 824 331 8	0.037 137 306 561 306 9
50	0.168 351 376 914 654 3	0.033 529 513 260 441 1	0.143 032 508 090 713 6

例 1.6 考虑以下 Lorenz 系统

$$\begin{cases}\dfrac{dx}{dt}=\sigma(y-x)\\ \dfrac{dy}{dt}=x(r-z)-y\\ \dfrac{dz}{dt}=xy-bz\end{cases} \tag{1.6}$$

其中

$$\sigma=10,\ r=28,\ b=\frac{8}{3}$$

取初值 $x_0=15.34$, $y_0=13.68$, $z_0=37.91$，利用四阶 Runge-Kute 法积分，积分步长 $h=0.04$，得关于变量 x 的长为 8500 的时间序列。为了减少瞬态的影响，去掉前 5000 个数据，剩下的 3500 个数据的时间历程如图 1.7 所示。

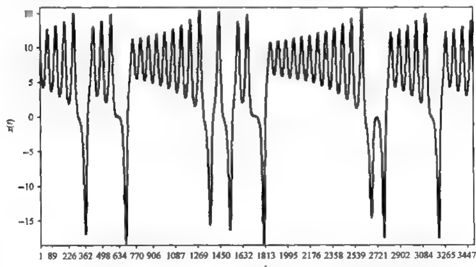


图 1.7 Lorenz 系统变量 x 的时间历程图

从图 1.7 中可以发现，随时间的演化，系统将出现不规则的振荡，看起来好像是随机的，实际上此时系统是混沌的。

4. 实际问题中的实测时间序列

实际问题中有大量的时间序列，如天文中的太阳黑子时间序列、水文中的径流量时间序列、气象中的降雨量时间序列、生理系统中的心电图时间序列和经济系统中的股票指数时间序列等。

例 1.7　股票指数时间序列。

图 1.8 是从 1999 年 1 月 4 日至 2003 年 12 月 31 日的上海证券市场综合股指时间序列，共 1196 个数据，可以看出股票指数时间序列表现为非常复杂的演化行为。

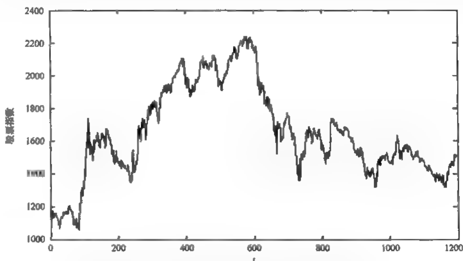


图 1.8　上海证券市场综合股指时间序列历程图

例 1.8　太阳黑子数时间序列。

图 1.9 是 1700 年以来的太阳黑子数时间序列，可以看出太阳黑子时间序列也表现为非常复杂的演化行为。

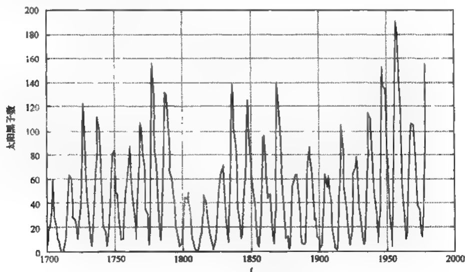


图 1.9　太阳黑子数时间序列历程图

从上面的时间序列例子可以看出，不同的时间序列具有不同的特性。例如，趋向于稳定状态、趋向于周期循环或出现看起来是毫无规律的混沌特征，具有相同特性的时间序列可能来自于用差分方程描述的非线性系统，也可能来自于由微分方程描述的系统。

实际问题中通过观测或实验手段获得的时间序列常常是比较复杂的，在不知道实际系统解析模型的情况下，分析观测时间序列的演变规律是掌握所研究系统动力学特性的重要手段。

1.3　研究非线性时间序列的意义

自 20 世纪 80 年代以来，有相当多的学者利用混沌、分形等非线性动力系统理论来研究实际问题中的复杂系统，并取得了不少研究成果。非线性动力系统理论是定量研究复杂系统的一种较好的方法，对于那些其演化过程不可预测但其不可预测性可以预测的复杂系统，即确定性系统内具有内在随机性的复杂系统，混沌理论成为一种有效的研究工具。利用非线性动力系统理论研究复杂系统需要建立非线性动力学模型，但实际问题中的复杂系统往往表现为多变量动态演化行为、多层次结构，由于获取的信息通常是不完备的，并具不确定性，且系统的结构参数和边界条件具有时变性和复杂性，因此往往难以直接建立起精确的解析形式的数学模型，常常依赖于通过观测或实验手段获取的时间序列进行分析。

如果通过观测或实验手段获得的时间序列是来自于线性系统，则时间序列可表示成不同幅度的正弦波的叠加，因此可用传统的线性方法和 Fourier 变换方法进行处理。但当获得的时间序列来自于非线性系统时，无论多么高阶次的线性模型都不可能对该非线性系统的行为做出恰当的解释；相反，一个低阶非线性模型或许能很好地刻画该动力系统，对这样的时间序列，非线性时间序列分析方法成为重要的研究工具。

经典的时间序列分析方法是把时间序列看成随机过程进行研究的，因而整个理论是以随机过程作为基础的。随着混沌现象的发现，人们发现在确定性系统内部也存在随机性，但这种复杂系统的本质特征往往不是随机因素而是非线性动力系统中的混沌因素造成的，对从这样的系统获取的时间序列，用随机过程方法分析显然是不合适的，而应当用混沌时间序列分析方法进行研究。

混沌运动与牛顿力学中运动的三种定常状态（静止、周期运动和准周期运动）不同，它是一种始终限于有限区域、轨道永不重复、性态复杂的运动，它具有对初值的极端敏感性、长期不可预测性及分形结构等特点，它的时间历程是非周期的或被认为是周期无穷大的、貌似随机的运动曲线。混沌运动不是随机运动，而是属于确定性运动的范畴，因为它是源于确定性系统，受到确定性激励，因而反映了确定性

系统的内在随机性。如果一个时间序列是由一个确定性系统通过观测或实验手段获得的，要考虑的是以下反问题：如何由时间序列来恢复并刻画原复杂系统？非线性时间序列分析就是研究如何通过复杂系统的观测或实验获得的时间序列来辨识和重构原复杂系统，分析原复杂系统的性质，刻画原复杂系统的特征，并对原复杂系统进行预测和控制。因此，研究非线性时间序列分析方法对解决实际问题中复杂系统的相关问题具有重要的意义。

第二章　单变量非线性时间序列分析

非线性时间序列分析方法的研究对象是未知数学模型或无法建立解析数学模型的复杂系统，研究过程中具有的信息是从复杂系统中通过观测或实验手段获得的单变量或多变量时间序列。研究目标就是寻求反映复杂系统本质特征的不变量，分析复杂系统的特征、内在变化规律及内部关系，最终解释、指导甚至控制实际问题中具体的复杂系统。

2.1　非线性时间序列分析流程

非线性时间序列分析可以看作是如图 2.1 所示的输入输出过程，输入的是没有任何背景知识的通过观测或实验手段获得的实际复杂系统的时间序列，输出的是所研究系统的某些特征。

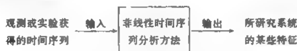


图 2.1　非线性时间序列分析的框架模型

为了具体说明如何从所研究的系统获得的非线性时间序列出发去研究系统的特征，下面通过两个例子来引出非线性时间序列分析的相空间重构思想。

例 2.1　考虑以下 Henon 映射

$$\begin{cases} x_{n+1} = 1 - 1.4x_n^2 + y_n \\ y_{n+1} = 0.3x_n \end{cases} \tag{2.1}$$

该系统虽然有两个状态变量 x_n, y_n，但由于

$$x_{n+1} = 1 - 1.4x_n^2 + 0.3x_{n-1} \tag{2.2}$$

表明该系统实际上只与状态变量 x_n 的前两个时刻的状态有关，因此，如果未知该系统的数学模型，且只观测到状态变量 x_n 的有关信息，我们也可以从状态变量 x_n 来建立原系统的模型。对状态变量 x_n 进行相空间重构，即令

$$\boldsymbol{x}_n = (x_n, x_{n-1}) \tag{2.3}$$

从而

$$x_{n+1} = G(\boldsymbol{x}_n) = G(x_n, x_{n-1}) = 1 - 1.4x_n^2 + 0.3x_{n-1} \tag{2.4}$$

也就是说，只要能找到函数G，就可以从观测时间序列$\{x_n\}$重构原来的系统。

例 2.2 考虑以下 Lorenz 系统

$$\begin{cases} \dfrac{dx}{dt} = \sigma(y - x) \\ \dfrac{dy}{dt} = x(r - z) - y \\ \dfrac{dz}{dt} = xy - bz \end{cases} \tag{2.5}$$

其中

$$\sigma = 10, r = 28, b = \frac{8}{3}$$

取初值$x_0=15.34$，$y_0=13.68$，$z_0=37.91$，利用四阶 Runge Kute 法积分，积分步长$h=0.04$，得关于状态变量x、y、z的长为 2000 的时间序列，从而可以画出如图 2.2 所示的 Lorenz 吸引子。另外，如果我们不知道该系统的数学模型，只观测到状态变量x的时间序列x_n，对状态变量x进行相空间重构，令

$$\boldsymbol{x}_n = (x_n, x_{n-4}, x_{n-8}) \tag{2.6}$$

从而可得如图 2.3 所示的 Lorenz 吸引子的重构图。由图 2.3 可以看出，实际 Lorenz 吸引子与重构的 Lorenz 吸引子非常类似。

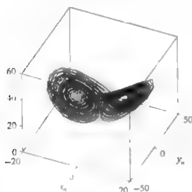


图 2.2 Lorenz 吸引子

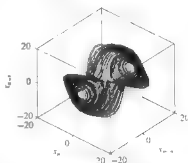


图 2.3 Lorenz 吸引子按(x_n, x_{n-4}, x_{n-8})重构的相图

上面两个例子中的系统都具有混沌特征，产生的时间序列是混沌时间序列，例子中的相空间重构就是混沌时间序列分析的基础。相空间重构的基本原理是 F. Takens 和 R. Mane[2]的延迟嵌入定理，它建立了观测信号系统时间波动和动力系统空间特征之间的桥梁。

单变量混沌时间序列分析的基本思想可以描述如下：

对未知数学模型的混沌动力系统，通过观测或实验手段可以获得单变量时间

序列$\{x_n\}$，做以下相空间重构

$$\boldsymbol{x}_n = (x_n, x_{n-\tau}, \cdots, x_{n-(m-1)\tau}) \tag{2.7}$$

从而形成 m 维状态空间，在重构的 m 维状态空间中可以建立数学模型

$$\boldsymbol{x}_{n+1} = \boldsymbol{G}(\boldsymbol{x}_n) \tag{2.8}$$

F. Takens 和 R. Mane 证明了只要适当选取 m 和 τ，原未知数学模型的混沌动力系统的几何特征与重构的 m 维状态空间的几何特征是等价的，它们具有相同的拓扑结构，这意味着原未知数学模型的混沌动力系统中的任何微分或拓扑不变量可以在重构的状态空间中计算，并且可以通过在重构的 m 维状态空间中建立数学模型对原未知数学模型的动力系统进行预测，进一步解释、分析、指导原未知数学模型的动力系统。这一过程可以用图 2.4 表示。

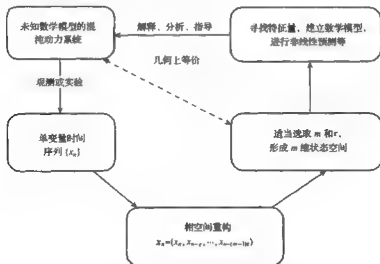


图 2.4　单变量混沌时间序列分析流程图

例 2.1 和例 2.2 中的时间序列就是系统的一个状态变量的观测值，没有观测噪声，而且已知原系统是混沌的，但实际问题中复杂系统通过观测或实验手段获得的时间序列往往没有这么理想，因此进行时间序列分析时还有许多工作需要做。

对实际问题中未知数学模型或无法建立解析数学模型的复杂系统，通过观测或实验手段获得的时间序列可能是不平稳的，可以通过适当的方法对时间序列进行平稳化处理，因为混沌时间序列分析方法主要是针对平稳时间序列的。同时，需要对通过观测或实验手段获得的时间序列进行非线性检验，如果时间序列是来自于线性系统，则可以通过线性方法对原系统进行建模；如果时间序列来自于非线性系统，需要继续检验其是来自于确定性系统还是来自于随机系统，对随机时间序列

可以通过统计方法研究，对来自确定性系统的时间序列，还需要进一步判断其是否具有混沌特征，因为对混沌时间序列只能进行短期分析，其长期演化行为无法正确描述。另外，通过观测或实验手段获得的时间序列往往具有噪声，在进行分析处理前需要进行降噪处理。在上述分析的基础上，可以对时间序列进行相空间重构，并在重构的相空间中建立数学模型来分析原系统的演化行为。

因此，通过观测或实验手段获得的时间序列的分析流程可以用图 2.5 表示。

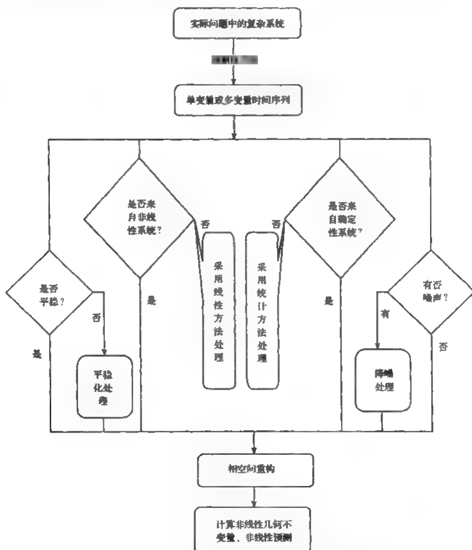


图 2.5　通过观测或实验手段获得时间序列的分析流程图

非线性时间序列分析的一个缺陷是从输出的结果即所研究系统的某些特征本身无法给出系统的实际意义解释，无法从这一过程获得系统的某些方面的改变对系统的影响，而通过建立系统数学模型研究系统行为的正向研究方法恰恰能做到这一点。但这并不妨碍对非线性时间序列分析方法的研究，因为在实际应用中经常是把正向的模型研究与反向的时间序列分析结合起来，根据实际系统的背景知识建立的正向模型可以给出系统的测量时间序列的特征，而通过对观测或实验手段获取的时间序列的分析可以检验和定量分析这些特征，同时非线性时间序列分析可以给出什么样的方程或模型适合所研究系统的相关信息，为建立系统动力学模型提供建议和参考。另外，也可以通过对已建立的系统动力学模型和非线性时间序列的预测正确程度的比较，分析已建立的系统动力学模型的精确性，为进一步修正系统动力学模型提供依据。

2.2 相空间重构

根据观测或实验手段获得的混沌时间序列去研究复杂系统这种方法的基础是相空间重构，它的基本思想是：由于混沌系统产生的轨迹经过一定时期的变化后，会最终做一种有规律的运动，产生一种规则的、有形的轨迹(混沌吸引子)。系统中任一分量的演化都是由与之相互作用的其他分量所决定的，因此这些相关分量的信息就隐含在任一分量的发展过程中，这样就可以从某一分量的时间序列中提取和恢复出动力系统原来的规律。为了重构一个等价的状态空间，只需考察一个分量，并将它在某些固定的时间延迟点上的观测量作为新维处理，即延迟值被看作是新的坐标，它们确定了某个多维状态空间中的一点。重复这一过程并测量相对于不同时间的各延迟量，就可以产生出许多这样的点，已经证明它可以将吸引子的许多性质保存下来，即用系统的一个观测量可以重构出原动力系统的模型，并可初步确定该系统的真实相空间的维数。由此可见，相空间重构就是从时间序列出发创建一个多维状态空间，它保持了原系统的许多几何不变量不变，这些几何不变量包括不动点的特征值、吸引子的分维数和轨线的 Lyapunov 指数等。相空间重构的依据是 F. Takens[1]和 R. Mane[2]的延迟嵌入定理。F. Takens 的论文对非数学专业的人来说很难读懂，R. Mane 的论文写得更抽象，但可通过下面的相空间重构过程来理解 F. Takens 和 R. Mane 的延迟嵌入定理。

设动力系统是由非线性差分方程

$$z_{n+1} = F(z_n) \tag{2.9}$$

表示的离散系统，或是由微分方程

$$\frac{dz(t)}{dt} = F(z(t)) \tag{2.10}$$

表示的连续系统。其中 z_n 或 $z(t)$ 是系统在时刻 n 或 t 的状态向量，$F(\cdot)$ 是向量值函数。时间序列 $\{x_n\}$ 是观测到的系统某一维输出，即

$$x_n = h(z_n) + \omega_n \tag{2.11}$$

或

$$x_n = x(t_0 + n\Delta t) = h[z(t_0 + n\Delta t)] + \omega_n \tag{2.12}$$

式中，$h(\cdot)$ 为多元数量值函数；ω_n 为在观测或者测量过程中由于技术手段不完善或者精度不够引起的测量噪声。

根据 F. Takens 的定理[1]，当 $\omega_n=0$ 时，观察到的时间序列 $\{x_n\}$ 以向量

$$\boldsymbol{x}_n = (x_n, x_{n-\tau}, \cdots, x_{n-(m-1)\tau}) \tag{2.13}$$

形成 m 维空间，只要 $m\geqslant 2d+1$，动力系统的几何结构可以完全打开，其中 d 是系统中吸引子的维数，τ 是正整数，称为延迟时间间隔。条件 $m\geqslant 2d+1$ 是动力系统重构的充分但不必要条件，获得动力系统重构的整数 m 叫做嵌入维数。状态空间中 $\boldsymbol{x}_n \to \boldsymbol{x}_{n+1}$ 的演化反映了未知动力系统 $z_n \to z_{n+1}$ 或 $z(t) \to z(t+1)$ 的演化，并且状态空间 R^m 中吸引子的几何特征与原动力系统的几何特征等价，这意味着原动力系统中任何微分或拓扑不变量可以在重构的状态空间中计算。F. Takens 的定理是在无噪声的情况下考虑的，即式(2.11)和式(2.12)中的 $\omega_n=0$，后来 T. Sauer 等把延迟嵌入定理推广到了具有噪声的情形[3]。

对一组长为 N 的实测时间序列 $\{x_n\}_{n=1}^{N}$，由式(2.13)可构造出 m 维状态向量

$$\boldsymbol{x}_n = (x_n, x_{n-\tau}, \cdots, x_{n-(m-1)\tau}) \in R^m, \quad n = N_0, N_0+1, \cdots, N \tag{2.14}$$

其中 $N_0=(m-1)\tau+1$，τ 是延迟时间间隔。在 R^m 中以 L_2 或 L_∞ 范数定义 $\boldsymbol{x}_i$ 到 $\boldsymbol{x}_j$ 的距离，即

$$\| \boldsymbol{x}_i - \boldsymbol{x}_j \|_2 = \sqrt{\sum_{l=0}^{m-1} (x_{i+l\tau} - x_{j+l\tau})^2} \tag{2.15}$$

或

$$\| \boldsymbol{x}_i - \boldsymbol{x}_j \|_\infty = \max_{0\leqslant l\leqslant m-1} |x_{i+l\tau} - x_{j+l\tau}| \tag{2.16}$$

以下在不特别说明的情况下，记 R^m 中的范数为 $\|\cdot\|$，但 L_∞ 范数在计算距离时实现较快。

为了能在重构的 R^m 空间中刻画原动力系统的性质，需正确地确定延迟时间间隔 τ 和嵌入维数 m。

2.2.1 延迟时间间隔的确定

由 F. Takens 的延迟嵌入定理知，在时间序列无限长、无噪声、无限精确的情况下，可以任意选择 τ，但实测时间序列是有限长的，且一般都有噪声污染和测量误差，只能根据经验来选择 τ。选择 τ 的基本思想是使 x_n 与 $x_{n+\tau}$ 具有某种程度的独

立但又不完全无关，以便它们能在重构的相空间中作为独立的坐标处理。如果 τ 太小，则 x_n 与 $x_{n+\tau}$ 的值充分靠近，以至于不能区分它们，从实际观点看不能提供两个独立的坐标，导致吸引子重构非常靠近相空间中的"对角线"，重构的相空间可能总是杂乱无规则的；如果 τ 太大，x_n 与 $x_{n+\tau}$ 可能会不相关，吸引子轨道会投影在两个完全不相关的方向上，不能反映相空间中轨线的真实演化规则。鉴于此，需要选择一个比较合适的延迟时间间隔 τ。目前，可以使用的方法非常多，但从计算复杂性和使用的简便性等角度看，比较常用的方法主要有自相关函数法和平均互信息法。

1. 自相关函数法

自相关函数的基本思想是要考察观测量 x 和 $x_{n+\tau}$ 与平均观测量的差之间的线性相关性，即如果假设

$$x_{n+\tau} - \overline{x} = C_L(\tau)(x_n - \overline{x}) \tag{2.17}$$

其中

$$\overline{x} = \frac{1}{N}\sum_{n=1}^{N} x_n \tag{2.18}$$

则使

$$\sum_{n=1}^{N}\left[x_{n+\tau} - \overline{x} - C_L(\tau)(x_n - \overline{x})\right]^2 \tag{2.19}$$

最小的 $C_L(\tau)$ 为

$$C_L(\tau) = \frac{\frac{1}{N}\sum_{n=1}^{N}(x_{n+\tau} - \overline{x})(x_n - \overline{x})}{\frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})^2} \tag{2.20}$$

称这样的 $C_L(\tau)$ 为线性自相关函数。取 $C_L(\tau)$ 第一次为零时的 τ 为延迟时间间隔，此时在平均意义下，x_n 和 $x_{n+\tau}$ 是线性无关的[4]。

自相关函数法是比较简单的寻找时间延迟 τ 的一种方法，但这种方法只考虑到时间序列中线性关系，至于非线性关系并不清楚，所以并不适合所有情况。特别当自相关函数变化十分缓慢时，选择 τ 会非常困难。

2. 平均互信息法

不同于自相关函数法，平均互信息法将非线性关系也考虑在内，这种方法的根据是可从事件 b_j 在集 B 中发生的概率中得到关于 a_i 在集 A 发生概率的信息。文献[5，6]给出了平均互信息的定义以及如何从时间序列计算平均互信息的方法。

设观测时间序列为 $\{x_n\}$，则在 n 和 $n+\tau$ 时刻观测量之间的平均互信息为

$$I(\tau) = \sum_{n=1}^{N} P(x_n, x_{n+\tau}) \log_2\left[\frac{P(x_n, x_{n+\tau})}{P(x_n)P(x_{n+\tau})}\right] \tag{2.21}$$

式中，$P(x_n)$，$P(x_{n+\tau})$，$P(x_n,x_{n+\tau})$为概率。

概率$P(x_n)$，$P(x_{n+\tau})$可以通过计算时间序列的直方图获得，联合概率$P(x_n,x_{n+\tau})$可以通过计算时间序列的二维直方图获得，利用计算机可以很方便地计算观测时间序列的平均互信息[5,6]。

文献[5]建议选择$I(\tau)$的第一个局部最小时的τ为延迟时间间隔，因为此时产生的冗余最小，产生了最大的独立性。与自相关函数法相比，平均互信息法考虑了非线性依赖性，但仍有其局限性，如有时可能无局部最小或对某些例子特别不合适[7,8]。

除了自相关函数法和平均互信息法外，选择τ的方法还有重构展开法[8]、高阶关联法[9]、通过分析整体和局部混沌吸引子行为获得优化延迟时间的填充因子法[10]等多种方法，这些方法都有各自的特点，但实际应用中用得较多的还是自相关函数法和平均互信息法。

2.2.2　嵌入维数的确定

关于嵌入维数，F. Takens[1]、R. Mane[2]、T. Sauer[3]等先后从理论上证明了当$m\geqslant 2d+1$时可获得一个吸引子的嵌入，其中d是吸引子的分形维数，但这只是一个充分条件，对观测时间序列选择m没有帮助。如果仅仅是计算关联维数，M. Z. Ding等[11]证明了对无噪声、无限长的时间序列，只要取m为大于关联维数d的最小整数即可，但对长度有限且具噪声的时间序列，m要比d大得多。如果m选得太小，则吸引子可能折叠以致在某些地方自相交，这样在相交区域的一个小邻域内可能会包含来自吸引子不同部分的点；如果m选得太大，理论上是可以的，但在实际应用中，随着m的增加会大大增加吸引子几何不变量（如关联维数、Lyapunov指数等）的计算工作量，且噪声和舍入误差的影响会大大增加。

在实际应用中常有以下方法选取嵌入维数m。

1. *试算法*

试算法是通过逐步增加计算过程中的嵌入维数，观察什么时候某些几何不变量（例如，关联积分、关联维数、Lyapunov指数等）停止变化的方法。从理论上讲，由于这些几何不变量是吸引子的几何性质，当m大于最小嵌入维数时，几何结构被完全打开，因此这些不变量与嵌入维数无关，取吸引子的几何不变量停止变化时的m为最小嵌入维数。这种方法的缺点是对数据要求较高（无噪声）、计算量大且比较主观。例如，P. Grassberger通过增加嵌入维数m，计算关联积分$C_N(r,m,\tau)$，取当关联积分$C_N(r,m,\tau)$不再变化时的m为嵌入维数[12]。

2. *虚假邻点法*

虚假邻点法建立在以下事实的基础上：选择太小的嵌入维数将导致那些在原

相空间中离得比较远的点会在重构的相空间中靠近。其基本思想是当嵌入维数从 m 变到 $m+1$ 时，考察轨线 x_n 的邻点中哪些是真实的邻点，哪些是虚假的邻点，当没有虚假邻点时，可以认为几何结构被完全打开[13]。

设 x_n 的最近邻点为 $x_{\eta(n)}$，即

$$\| x_{\eta(n)} - x_n \|_2^{(m)} = \min_{j=N_0,\cdots,N, j\neq n} \| x_n - x_j \|_2 = \sqrt{\sum_{l=1}^{m-1} (x_{\eta(n)+l\tau} - x_{n+l\tau})^2} \tag{2.22}$$

当嵌入维数从 m 增加到 $m+1$ 时，它们之间的距离变为

$$\| x_{\eta(n)} - x_n \|_2^{(m+1)} = \sqrt{\sum_{l=0}^{m} (x_{\eta(n)+l\tau} - x_{n+l\tau})^2} = \sqrt{[\| x_{\eta(n)} - x_n \|_2^{(m)}]^2 + [x_{\eta(n)+m\tau} - x_{n+m\tau}]^2} \tag{2.23}$$

若 $\| x_{\eta(n)} - x_n \|_2^{(m+1)}$ 比 $\| x_{\eta(n)} - x_n \|_2^{(m)}$ 大很多，可以认为是由于高维吸引子中两个不相邻的点在投影到低维轨线上时变成相邻的两点造成的，因此这样的邻点是虚假的。具体地说，当

$$\sqrt{\frac{[\| x_{\eta(n)} - x_n \|_2^{(m+1)}]^2 - [\| x_{\eta(n)} - x_n \|_2^{(m)}]^2}{[\| x_{\eta(n)} - x_n \|_2^{(m+1)}]^2}} = \frac{|x_{\eta(n)+m\tau} - x_{n+m\tau}|}{\| x_{\eta(n)} - x_n \|_2^{(m+1)}} \geqslant R_{tol} \tag{2.24}$$

时，$x_{\eta(n)}$ 为 x_n 的虚假最近邻点，其中 R_{tol} 是某一个阈值，文献[13]建议 $R_{tol} \geqslant 10$，而文献[14]建议 R_{tol} 可取 15 左右。

观测时间序列通常具有噪声且长度有限，所以仅仅用上面的标准判别虚假最近邻点会不正确，为此，M. B. Kennel 等提出了增加以下标准，即当

$$\frac{\| x_{\eta(n)} - x_n \|_2^{(m+1)}}{\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[x_k - \frac{1}{N}\sum_{k=1}^{N} x_k\right]^2}} > A_{tol} \tag{2.25}$$

时，$x_{\eta(n)}$ 为 x_n 的虚假最近邻点，其中 A_{tol} 是阈值，文献[13]建议 $A_{tol}=2$。

对式(2.14)中的每个 x_n，按式(2.22)找出其最近邻点，并按照式(2.24)、式(2.25)的标准判别最近邻点是否是虚假的，从而可以计算出虚假最近邻点的比例。让 m 从 1 开始增加，计算每个 m 时的虚假最近邻点的比例，直到虚假最近邻点的比例小于 5%或虚假最近邻点不再随着 m 的增加而减少时，可以认为吸引子几何结构完全打开，此时的 m 为嵌入维数。

从几何观点看，这是一种较好的方法，但在判断虚假邻点时阈值的不同选取会导致不同的结果，因此具有较大的主观性。从式(2.24)看，阈值应当由时间序列的方差确定，因此，不同的时间序列有不同的阈值，这意味着要给出一个合适和合理

的阈值是非常困难的甚至是不可能的，为此，文献[15]提出了改进的虚假邻点算法。

3. 改进的虚假邻点法

对式(2.14)中的 x_n，类似于虚假最近邻点法的思想，文献[15]定义

$$a(n,m)=\frac{\| x_{\eta(n)}-x_n \|_{\infty}^{(m+1)}}{\| x_{\eta(n)}-x_n \|_{\infty}^{(m)}} \tag{2.26}$$

这里用的是 L_{∞} 范数，记所有 $a(n,m)$ 关于 n 的均值为

$$E(m)=\frac{1}{N-N_0+1}\sum_{n=N_0}^{N}a(n,m) \tag{2.27}$$

由于式(2.14)的相空间重构方式与文献[15]中有所差别，式(2.14)是向后重构，而文献[15]中的重构是向前重构，所以式(2.27)与文献[15]有所不同，但本质上是一样的。

$E(m)$ 只依赖于嵌入维数 m 和延迟时间间隔 τ，为了研究嵌入维数从 m 变为 $m+1$ 时相空间的变化情况，定义

$$E_1(m)=\frac{E(m+1)}{E(m)} \tag{2.28}$$

如果当 m 大于某个 m_0 时，$E_1(m)$ 停止变化，则 m_0+1 就是重构相空间的最小嵌入维数[15]。

对来自于随机系统的时间序列，原则上，随着 m 的增加，$E_1(m)$ 将永远不会达到饱和值，但在实际计算中，当 m 充分大时，很难分辨 $E_1(m)$ 是在缓慢增加还是停止变化。事实上，由于获得的观测时间序列是有限长的，可能会出现虽然是随机时间序列，但 $E_1(m)$ 会在某一 m 处停止变化。为了解决这一问题，文献[15]定义了量

$$E^*(m)=\frac{1}{N-N_0+1}\sum_{n=N_0}^{N}\left| x_{\eta(n)-m\tau}-x_{n-m\tau} \right| \tag{2.29}$$

记

$$E_2(m)=\frac{E^*(m+1)}{E^*(m)} \tag{2.30}$$

对随机时间序列，由于将来的值与过去的值独立，对任何 m，$E_2(m)$ 将等于 1，而对确定性系统的观测时间序列，$E_2(m)$ 显然与 m 有关，对所有 m，$E_2(m)$ 不可能为常数。因此，文献[15]建议可以通过同时计算 $E_1(m)$ 和 $E_2(m)$ 来确定时间序列相空间重构的最小嵌入维数，同时也区分了时间序列是来自于确定性系统还是随机系统。

除了上述方法外，确定嵌入维数的方法还有关联积分法[16]、奇异值分解法[17]以及上述各种的改进方法。

2.3 几何不变量的计算

对实际问题中具有混沌特征的复杂系统，通常可以通过一些几何不变量(例如，分形维数、Kolmogorov 熵和 Lyapunov 指数等)来定量刻画该系统的复杂程度，因此如何从观测或实验手段获得的时间序列去计算复杂系统的这些几何不变量是实际应用时所关心的问题。复杂系统的吸引子具有分数维数、正的 Kolmogorov 熵以及正的最大 Lyapunov 指数是判断系统是混沌的必要条件，但并不充分。分形维数的定义有很多种，从时间序列的角度看，关联维数是易于计算的一种分形维数。1983 年，P. Grassberger 和 I. Procaccia 提出了从时间序列中计算关联维数的 G-P 算法[12, 18]以及计算 Kolmogorov 熵的公式[19]。关联维数是系统复杂性程度的一种很好的度量。一般认为，大于关联维数的下一个整数是刻画系统所需的独立变量的个数，这为从时间序列恢复原复杂系统确定了一个框架。Lyapunov 指数度量了复杂系统的预测性，定量地刻画了初始靠近的状态空间轨线的指数发散率，从时间序列中计算最大 Lyapunov 指数的最早工作是于 1985 年由 A. Wolf 等提出的轨线法[20]。

2.3.1 关联维数

相空间的最小嵌入维数 m 是系统复杂性的一个界，这种拓扑维数在一般情况下只是吸引子占据空间数量的一个非常粗糙的描述，为了给出更严格的描述，文献[21～23]引入了几种不同的维数概念，其中关联维数是最常用的一种维数度量，尤其对通过观测或实验获得的时间序列，可由 P. Grassberger 和 I. Procaccia 给出的 G-P 算法直接计算[12, 18]。关联维数度量了所研究系统的复杂性，刻画了相空间中点的分布。在确定性系统中，关联维数就是生成相应复杂系统所必需的独立变量的个数，规则的确定性系统有整数关联维数，而混沌系统有非整数关联维数，大于此关联维数的下一个整数就是系统的独立变量的个数。但是某些关联的随机过程中也有非整数维数[24]，文献[25]证明了非整数关联维数不能成为混沌判断的充分条件，因为分形的 Browian 运动，虽然不是混沌的，但也有非整数关联维数。

G-P 算法描述如下：在按式(2.14)重构的 m 维空间中，$\sum_{j\neq i}H(r-\|x_i-x_j\|)$ 表示除 x_i 本身外到 x_i 的距离小于 r 的 x_j 的点数，其中 $H(\cdot)$ 是 Heavside 函数，即当 $x\leqslant 0$ 时，$H(x)=0$，当 $x>0$ 时，$H(x)=1$。定义按上述方式重构的时间序列 $\{x_n\}_{n=1}^{N}$ 的关联积分为

$$C_N(r,\tau,m)=\frac{2}{(N-N_0)(N-N_0+1)}\sum_{i=N_0}^{N-1}\sum_{j=i+1}^{N}H(r-\|x_i-x_j\|) \quad (2.31)$$

$C_N(r,\tau,m)$描述了距离小于 r 的对点数的分布情况。如果在 r 的某一区间段内，有

$$C_N(r,\tau,m) \infty r^d \tag{2.32}$$

则称 d 是关联维数，这样定义的 d 就是近似刻画产生时间序列$\{x_n\}_{n=1}^N$的复杂系统复杂程度的某种维数。对由观测或实验手段获得的时间序列，使式(2.32)成立的 r 的某一区间段可通过对双对数 $\ln C_N(r,\tau,m)$-$\ln r$ 图观测的办法或通过文献[26]给出的方法寻找。

G-P 算法是计算关联维数的一种经典方法，后来又有不少文献讨论了 G-P 算法的改进和其他计算关联维数的方法[27~34]。

由观测或实验手段获得的时间序列对应混沌系统的关联维数就是双对数 $\ln C_N(r,\tau,m)$-$\ln r$ 图的线性部分的斜率，由于连续采样点之间的相关性，会造成"肩峰"效应，为了消除这种"肩峰"效应，文献[27]对关联积分作了以下修改，即定义

$$C_N(r,\tau,m,w) = \frac{2}{(N-N_0-w+1)(N-N_0-w+2)} \sum_{i=N_0}^{N-w} \sum_{j=i+w}^{N} H(r - \|x_i - x_j\|) \tag{2.33}$$

式(2.33)中忽略了嵌入空间中非常靠近的点对关联积分的贡献，只考虑满足 $|i-j|\geqslant w$ 的点，其中 $w>\tau(2/N)^{2/m}$，特别取 $w=\tau$ 即可。文献[27]指出 w 的严格选取并不重要。

在实际应用中，关联维数的计算是非常费时的，为了改善前面算法的计算效率，文献[35]提出了一种修改算法。在关联积分中，较短的距离起着更有意义的作用，因此选取 r 的截断距离 r_0，构造宽度为 r_0 的盒子网格，只计算在相同的或相邻的盒子中距离小于 r_0 的对点的距离，这样可以把计算时间从 $O(N^2)$ 减到 $O(N\log_2 N)$。文献[35]给出了 r_0 的一个优化选择

$$r_0 = R(2/N)^{1/d} \tag{2.34}$$

式中，R 为时间序列的有效半径；N 为时间序列的长度；d 为关联维数的粗略估计。

当关联维数的值较大时，所需的时间序列要求较长[36]，而实际问题中，观测或实验获得的时间序列一般都比较短，为解决这一问题，文献[30]利用极大似然法估计关联维数的方法就有对数据要求相对较少的优点，其计算公式为

$$d = \left[\frac{\sum_{i>j} H(r_0 - \|x_i - x_j\|)\ln(r_0/\|x_i - x_j\|)}{\sum_{i>j} H(r_0 - \|x_i - x_j\|)}\right]^{-1} \tag{2.35}$$

用式(2.35)估计 d 的近似方差为

$$\mathrm{Var}(d) = \frac{2d^2}{p(N-N_0)(N-N_0+1)}[2q(N-2)+1] \tag{2.36}$$

其中

$$p = p\{ \| x_i - x_j \| < r_0 \} \tag{2.37}$$

式中，q 为 $1-d\ln \frac{r_0}{\| x_i - x_j \|}$ 和 $1-d\ln \frac{r_0}{\| x_i - x_k \|}$ 之间的相关系数，$\| x_i - x_j \| < r_0$，$\| x_i - x_k \| < r_0$。

这一方法的关键在于 r_0 的选取，既不能太小，也不能太大。当 r_0 较小时，可计算的点减少，则估计变得不可靠；当 r_0 较大时，虽然可以产生较稳定的结果，但由于 $C(r)=(r/r_0)^d(0<r\leqslant r_0)$ 的距离分布方差使结果产生较大的偏差，文献[30]详细给出了利用 χ^2 -测试确定 r_0 的一种方法。

虽然许多研究者提供了关联维数估计算法的精细的数学分析性质，但这些方法如何在有限长的、可能被噪声污染的、非平稳的观测或实验时间序列上实现仍有许多问题需要解决。

2.3.2 Kolmogorov 熵和 Renyi 熵

考察 m 维动力系统在吸引子上的轨道 $x(t)=(x_1(t),x_2(t),\cdots,x_m(t))$，把相空间分割成为边长为 r 的 m 维超立方体，每隔时间 Δt 去测量系统的状态，设 $p(i_1,i_2,\cdots,i_m)$ 是 $x(r=\Delta t)$ 在超立方体 i_1 中，$x(t=2\Delta t)$ 在超立方体 i_2 中，…，$x(t=m\Delta t)$ 在超立方体 i_m 中的联合概率分布，则 Kolmogorov 熵为

$$K_1 = -\lim_{\Delta t\to 0}\lim_{r\to 0}\lim_{m\to\infty}\frac{1}{m\Delta t}\sum_{i_1 i_2\cdots i_m} p(i_1,i_2,\cdots,i_m)\log_2 p(i_1,i_2,\cdots,i_m) \tag{2.38}$$

而 q 阶 Renyi 熵定义为

$$K_q = -\lim_{\Delta t\to 0}\lim_{r\to 0}\lim_{m\to\infty}\frac{1}{m\Delta t}\frac{1}{q-1}\log_2 \sum_{i_1 i_2\cdots i_m} p^q(i_1,i_2,\cdots,i_m) \tag{2.39}$$

显然有

$$\lim_{q\to 1} K_q = K_1 \tag{2.40}$$

若类似于关联积分，定义下面的 q 阶广义关联积分

$$C_N^q(r,\tau,m) = \left\{\frac{1}{(N-N_0+1)}\sum_{i=N_0}^{N}\left[\frac{1}{N-N_0}\sum_{j=N_0,j\neq i}^{N} H(r-\| x_i - x_j \|)\right]^{q-1}\right\}^{\frac{1}{q-1}} \tag{2.41}$$

则 q 阶 Renyi 熵可表示为[37]

$$K_q = -\lim_{\Delta t\to 0}\lim_{r\to 0}\lim_{m\to\infty}\frac{1}{m\Delta t}\log_2 C_N^q(r,\tau,m) \tag{2.42}$$

P. Grassberger 和 I. Procaccia 证明了 K_q 关于 q 是单调减少的[19]，因而 $K_1\geqslant K_2$。特别 $q=2$ 时，$C_N^2(r,\tau,m)$ 就是关联积分，因此可由关联积分计算二阶 Renyi 熵 K_2，它可作为 Kolmogorov 熵 K_1 的一个下界。

从观测时间序列计算熵一般是非常困难的，因为与计算关联维数等相比，计算熵需要更长的时间序列，有关熵的计算方面的更多工作可见文献[38～40]。

2.3.3 Lyapunov 指数

混沌动力系统有对初始条件的敏感依赖性，这意味着在状态空间中无限靠近的两个向量会产生两条指数分离的轨线，设 D 维混沌动力系统为

$$z_{n+1} = F(z_n) \tag{2.43}$$

设 $z_n^{(1)}$ 和 $z_n^{(2)}$ 为 D 维状态空间中两个非常靠近的状态，则它们的分离程度的时间演化为

$$\begin{aligned} z_{n+1}^{(1)} - z_{n+1}^{(2)} &= F(z_n^{(1)}) - F(z_n^{(2)}) \\ &= J_n(z_{n+1}^{(1)} - z_{n+1}^{(2)}) + O(\| z_n^{(1)} - z_n^{(2)} \|^2) \end{aligned} \tag{2.44}$$

其中 $F(z_n^{(1)})$在 $z_n^{(2)}$ 处进行 Taylor 展开，$J_n = J_n(z_n^{(2)})$是 F 在 $z_n^{(2)}$ 处的 $D \times D$ 阶 Jacobi 矩阵。

当扰动$\delta_n = z_n^{(1)} - z_n^{(2)}$ 给定时，可以计算一次迭代后它的变化情况。设 e_i 为 J_n 的特征向量，对应的特征值为 Λ_i，把δ_n以系数 β_i 分解为这些特征向量的和，即 $\delta_n = \sum \beta_i e_i$，则$\delta_{n+1} = \sum \beta_i \Lambda_i e_i$ 。特别地，如果δ_n平行于某个特征向量 e_i，则它以 Λ_i 为因子进行伸展或压缩，这样可找到 D 个不同的局部伸展因子，把状态空间分解为 D 个线性子空间。显然，对状态空间中任意其他点，可以找到 Jacobi 矩阵的不同特征值和特征向量，因为 Jacobi 矩阵本身是依赖于位置的。

为了从整体上刻画系统，引入不同局部因子的一个适当的平均，即 Lyapunov 指数，定义为

$$\lambda_i = \lim_{N \to \infty} \frac{1}{N} \ln |\Lambda_i^{(N)}|, \quad (i = 1, 2, \cdots, D) \tag{2.45}$$

其中 $\Lambda_i^{(N)}$ 定义为

$$\prod_{n=1}^{N} J_n u_i^{(N)} = \Lambda_i^{(N)} u_i^{(N)} \tag{2.46}$$

D 个不同的 λ_i 称为 Lyapunov 指数谱，其中最大的一个称为最大 Lyapunov 指数。如果最大 Lyapunov 指数是正的，意味着相邻的轨线按指数发散，即系统是混沌的。

对观测获得的混沌时间序列，最大 Lyapunov 指数可由 A. Wolf 等提出的轨线法计算[20]，而 Lyapunov 指数谱可由 J. P. Eckmann 和 R. Brown 等提出的矩阵法计算[41,42]。

1. 观测时间序列的最大 Lyapunov 指数的轨线计算法

1985 年，A. Wolf 等在文献[20]中提出了计算由观测获得的混沌时间序列的

最大 Lyapunov 指数的一种方法。

步骤 1：对时间序列$\{x_n\}_{n=1}^{N}$，做以下延迟嵌入

$$\boldsymbol{x}_n = (x_n, x_{n+\tau}, \cdots, x_{n+(m-1)\tau}) \in R^m,\ n = 1,2,\cdots,N-(m-1)\tau \quad (2.47)$$

步骤 2：以初始点$\boldsymbol{x}_1=(x_1,x_{1+\tau},\cdots,x_{1+(m-1)\tau})$为基点，在点集$\{\boldsymbol{x}_n\}$，$n=1,2,\cdots,N-(m-1)\tau$中除$\boldsymbol{x}_1$外的其余点中选取一个与$\boldsymbol{x}_1$最近的点$\boldsymbol{x}_{n_1}$作为端点，记$L_1=\|\boldsymbol{x}_1-\boldsymbol{x}_{n_1}\|$。

步骤 3：取一适当的时间步长或演化时间T_1，此时基点变为$\boldsymbol{x}_{1+T_1}$，端点变为$\boldsymbol{x}_{n_1+T_1}$，记$L_1'=\|\boldsymbol{x}_{1+T_1}-\boldsymbol{x}_{n_1+T_1}\|$，则经过$T_1$时间演化后，可得指数增长率为

$$\lambda_1 = \frac{1}{T_1}\log_2\frac{L_1'}{L_1} \quad (2.48)$$

演化时间T_1不能太长，否则有可能因为吸引子整体结构上的折叠、突变而产生对 Lyapunov 指数的错误估计。

步骤 4：在点集$\{\boldsymbol{x}_n\}$，$n=1,2,\cdots,N-(m-1)\tau$中除$\boldsymbol{x}_{1+T_1}$外的其余点中选取一个与$\boldsymbol{x}_{1+T_1}$最近的点$\boldsymbol{x}_{n_2}$作为新的端点进行向量替换，且使向量$\boldsymbol{x}_{n_1+T_1}\boldsymbol{x}_{1+T_1}$与$\boldsymbol{x}_{n_2}\boldsymbol{x}_{1+T_1}$之间的夹角$\theta_1$尽可能小，记$L_2=\|\boldsymbol{x}_{1+T_1}-\boldsymbol{x}_{n_2}\|$。

步骤 5：取一个新的时间步长或演化时间T_2，此时基点的演化为$\boldsymbol{x}_{1+T_1+T_2}$，端点的演化为$\boldsymbol{x}_{n_2+T_2}$，记$L_2'=\|\boldsymbol{x}_{1+T_1+T_2}-\boldsymbol{x}_{n_2+T_2}\|$，则经过$T_2$时间演化后，可得指数增长率为

$$\lambda_2 = \frac{1}{T_2}\log_2\frac{L_2'}{L_2} \quad (2.49)$$

在点集$\{\boldsymbol{x}_n\}$，$n=1,2,\cdots,N-(m-1)\tau$中除$\boldsymbol{x}_{1+T_1+T_2}$外的其余点中选取一个与$\boldsymbol{x}_{1+T_1+T_2}$最近的点$\boldsymbol{x}_{n_3}$作为新的端点进行向量的替换，且使向量$\boldsymbol{x}_{n_2+T_2}\boldsymbol{x}_{1+T_1+T_2}$与$\boldsymbol{x}_{n_3}\boldsymbol{x}_{1+T_1+T_2}$之间的夹角$\theta_2$尽可能小，记$L_3=\|\boldsymbol{x}_{1+T_1+T_2}-\boldsymbol{x}_{n_3}\|$。

步骤 6：重复上述过程，直到时间步长或演化时间到达点集$\{\boldsymbol{x}_n\}$，$n=1,2,\cdots,N-(m-1)\tau$的终点，记演化的总步数为M，则取指数增长率λ_i的平均值作为最大 Lyapunov 指数的估计值，即

$$\lambda = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{T_i}\log_2\frac{L_i'}{L_i} \quad (2.50)$$

步骤 7：增加嵌入维数m，并重复上述计算步骤，直到指数的估计值随m保持平稳为止，此时得到的计算结果即为所求的最大 Lyapunov 指数。

上述过程可用图 2.6 表示。

综上可知，为了得到一个好的最大 Lyapunov 指数的估计值，必须要求观测时间序列具有足够的长度。最大 Lyapunov 指数的轨线算法的一种改进可见文献[43]。

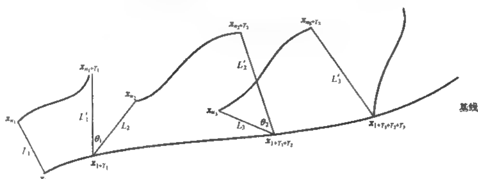


图 2.6　最大 Lyapunov 指数的轨线计算方法

2. Lyapunov 指数谱的矩阵计算法

Lyapunov 指数谱的矩阵计算法最初由 J. P. Eckmann 和 R. Brown 等在文献[41,42]中提出，对一组长为 N 的实测时间序列 $\{x_n\}_{n=1}^{N}$，在按式(2.14)重构的 m 维空间中，定义从向量 $\boldsymbol{x}_n$ 演化到 $\boldsymbol{x}_{n+v}$ 的演化算子为 $\boldsymbol{G}=(G_1,G_2,\cdots,G_m)^{\mathrm{T}}$，即

$$\boldsymbol{x}_{n+v}^{\mathrm{T}}=\boldsymbol{G}(\boldsymbol{x}_n^{\mathrm{T}}) \tag{2.51}$$

写成分量形式为

$$x_{n+v-(l-1)\tau}=G_l(x_n,x_{n-\tau},\cdots,x_{n-(m-1)\tau}),\ l=1,2,\cdots,m \tag{2.52}$$

其中自然数 v 是演化时间间隔。设 $\boldsymbol{x}_n$ 的最近邻点为 $\boldsymbol{x}_{\eta(n)}$，则由一阶 Taylor 展开可得

$$\begin{aligned}\boldsymbol{x}_{\eta(n)+v}^{\mathrm{T}}-\boldsymbol{x}_{n+v}^{\mathrm{T}}&=\boldsymbol{G}(\boldsymbol{x}_{\eta(n)}^{\mathrm{T}})-\boldsymbol{G}(\boldsymbol{x}_n^{\mathrm{T}})\\&=\boldsymbol{G}(\boldsymbol{x}_n^{\mathrm{T}}+(\boldsymbol{x}_{\eta(n)}^{\mathrm{T}}-\boldsymbol{x}_n^{\mathrm{T}}))-\boldsymbol{G}(\boldsymbol{x}_n^{\mathrm{T}})\\&\approx \boldsymbol{DG}^{(n)}(\boldsymbol{x}_{\eta(n)}-\boldsymbol{x}_n)^{\mathrm{T}}\end{aligned} \tag{2.53}$$

其中 Jacobi 矩阵 $\boldsymbol{DG}^{(n)}$ 为

$$\boldsymbol{DG}^{(n)}=\begin{bmatrix}\dfrac{\partial G_1}{\partial x_n} & \dfrac{\partial G_1}{\partial x_{n-\tau}} & \cdots & \dfrac{\partial G_1}{\partial x_{n-(m-1)\tau}}\\ \dfrac{\partial G_2}{\partial x_n} & \dfrac{\partial G_2}{\partial x_{n-\tau}} & \cdots & \dfrac{\partial G_2}{\partial x_{n-(m-1)\tau}}\\ \vdots & \vdots & \vdots & \vdots\\ \dfrac{\partial G_m}{\partial x_n} & \dfrac{\partial G_m}{\partial x_{n-\tau}} & \cdots & \dfrac{\partial G_m}{\partial x_{n-(m-1)\tau}}\end{bmatrix} \tag{2.54}$$

则由 Oseledec 定理，动力系统(2.51)的所有 Lyapunov 指数为矩阵

$$\lim_{K\to\infty}\left[(\boldsymbol{DG}^{(N_0+K)}\boldsymbol{DG}^{(N_0+K-1)}\cdots\boldsymbol{DG}^{(N_0)})^{\mathrm{T}}(\boldsymbol{DG}^{(N_0+K)}\boldsymbol{DG}^{(N_0+K-1)}\cdots\boldsymbol{DG}^{(N_0)})\right]^{\frac{1}{2K}} \tag{2.55}$$

的特征值的对数。

对已知的动力系统，$DG^{(n)}$可由式(2.54)确定，而对由观测或实验获得的时间序列描述的未知动力系统，可按下面的方法来确定$DG^{(n)}$。

当演化时间间隔v与延迟时间间隔τ一致时，由式(2.52)、式(2.54)可得

$$DG^{(n)}=\begin{bmatrix}0 & 1 & \cdots & 0\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ a_1^{(n)} & a_2^{(n)} & \cdots & a_m^{(n)}\end{bmatrix} \tag{2.56}$$

式中，$a_l^{(n)}=\dfrac{\partial G_m}{\partial x_{n-(l-1)\tau}}$，$l=1,2,\cdots,m$为待定系数。

对给定的$r>0$，若x_n的r邻域内有$L(n,r)$个点，即$\|x_{n_k}-x_n\|\leqslant r$，$k=1,2,\cdots,L(n,r)$，则当$L(n,r)\geqslant m$时，可由以下最小二乘法确定$a_1^{(n)},a_2^{(n)},\cdots,a_m^{(n)}$，即找使

$$\begin{aligned}&f(a_1^{(n)},a_2^{(n)},\cdots,a_m^{(n)})\\ &=\sum_{k=1}^{L(n,r)}\Big[\sum_{l=1}^{m}a_l^{(n)}(x_{n_k-(l-1)\tau}-x_{n-(l-1)\tau})-(x_{n_k-m\tau}-x_{n-m\tau})\Big]^2\end{aligned} \tag{2.57}$$

取最小值的$a_1^{(n)},a_2^{(n)},\cdots,a_m^{(n)}$，此时式(2.53)的近似程度最好。

当演化时间间隔v与延迟时间间隔τ不一致时

$$DG^{(n)}=\begin{bmatrix}a_{11}^{(n)} & a_{12}^{(n)} & \cdots & a_{1m}^{(n)}\\ a_{21}^{(n)} & a_{22}^{(n)} & \cdots & a_{2m}^{(n)}\\ \vdots & \vdots & \vdots & \vdots\\ a_{m1}^{(n)} & a_{m2}^{(n)} & \cdots & a_{mm}^{(n)}\end{bmatrix} \tag{2.58}$$

中所有系数可由最小二乘法逐行来确定。

Lyapunov 指数谱矩阵算法的改进和快速算法可见文献[44～50]。

2.4　观测时间序列平稳性的检验

无论是传统的线性时间序列分析方法还是最近越来越受到重视的非线性时间序列分析方法，都必须假设时间序列的某种平稳性。对于通过观测或实验获得的时间序列，如何检验它是否具有某种平稳性是首先要解决的问题。

时间序列$\{x_n\}_{n=1}^{\infty}$称为平稳的，若对所有k及连续函数$h:R^{k+1}\to R$，都有

$$\bar{h}=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}h(x_n,x_{n+1},\cdots,x_{n+k}) \tag{2.59}$$

存在[51]。但是通过观测或实验获得的时间序列几乎不可能满足这个标准，为此文献[52]提出了n阶"弱平稳性"概念，即n阶矩关于时间不变。文献[53]说明二阶平稳(即时域中的均值、方差及频域中的能量谱在时间段内不变)，再加上正态假设足

以产生完全的平稳性。

设 $\{x_n\}_{n=1}^{N}$ 为一时间序列，把它分成若干长为 l 的邻近的数据段，记第 i 段为 $s_i^l=\{x_{(i-1)l+1},\cdots,x_{il}\}\left(i=1,2,\cdots,\frac{N}{l}\right)$。通常的方法是计算每一数据段 s_i^l 的统计量 γ_i（可以是均值、方差，也可以是关联维数、Lyapunov 指数等），如果序列 $\{\gamma_i\}$ 是近似的常数序列，可以认为这一时间序列是平稳的，但这种做法依赖于对统计量 γ_i 的性质的了解及它的概率分布。如果以均值和方差作为统计量，其优点是对数据段长度 l 的要求不高，一段 1000 点的时间序列足以进行可靠的均值和方差估计，且对线性时间序列的分析比较有效[54]。但由于没有考虑非线性性，对非线性时间序列就不太合适，虽然也有文献这样做了，但对于像心电图(ECG)、脑电图(EEG)这样具有复杂关系的非线性时间序列，即使在二阶矩不变的一个时间段内也可能是非平稳的。因此，出现用关联维数、Lyapunov 指数等代替均值、方差作为统计量的方法[55]，虽然能反映非线性性，但出现的问题是时间段的选取问题，因为太短的时间段无法计算关联维数、Lyapunov 指数，而太长的时间段，可能就是非平稳的，因此也无法计算关联维数、Lyapunov 指数。

也有不用比较连续统计量的方法检验时间序列平稳性的方法，如文献[56]的利用空间时间分离图的关联维数估计法、文献[57]的统计方法和文献[58, 59]的递归图法。递归图法检验时间序列平稳性的基本思想是画出相空间中每个开始点的最近邻点的坐标，如果形式对任一初始点都类似，则系统是平稳的，这种方法要用到多维模式识别、聚类分析等方法。

文献[60]提出一种新的方法是建立在时间序列不同部分本身间的相似性而不是从时间序列由局部平均导出参数的相似性的基础上，这种方法特别适合非平稳性是由吸引子形状变化造成的而此时动力系统不变量保持不变的情况。在文献[60]中定义了一对时间段 s_i^l,s_j^l 的统计量 $\gamma_{ij}=\gamma(s_i^l,s_j^l)$，原则上，$\gamma(s_i^l,s_j^l)$ 可以是任何对 s_i^l 和 s_j^l 差异敏感的统计量。对实际应用而言，γ_{ij}的严格理论定义并不重要，重要的是对相对较短的 s_i^l,s_j^l 获得稳定和可靠的估计，满足这一标准的一个统计量是一个简单非线性预测算法的误差（正交预测误差）。具局部常数近似的预测对几百或更少的点产生稳定的结果，而全局非线性预测可对更少的点实现。从理论上讲更吸引人的是文献[61]定义的正交关联积分，但为了获得稳定的结果，它要求较长的时间段和较低的噪声。

设 $s_i^l=\{x_{(i-1)l+1},x_{(i-1)l+2},\cdots,x_{il}\}$，$s_j^l=\{x_{(j-1)l+1},x_{(j-1)l+2},\cdots,x_{jl}\}$，$m$ 为最小嵌入维数，记

$$s_{i,n}^l=\{x_{(i-1)l+n-m+1},x_{(i-1)l+n-m+2},\cdots,x_{(i-1)l+n}\},\ n=m,m+1,\cdots,l-1 \tag{2.60}$$

$$s_{j,n}^l=\{x_{(j-1)l+n-m+1},x_{(j-1)l+n-m+2},\cdots,x_{(j-1)l+n}\},\ n=m,m+1,\cdots,l-1 \tag{2.61}$$

固定 ϵ 及 $s_{j,n}^{i}$，利用 $s_{i,n}^{i}$ 作为数据基估计下一步预测，即

$$x_{(j-1)l+n+1}^{*} = \frac{1}{|u_{\epsilon}(s_{j,n})|} \sum_{s_{i,p} \in u_{\epsilon}(s_{j,n})} x_{(i-1)l+p+1} \tag{2.62}$$

其中

$$u_{\epsilon}(s_{j,n}) = \{s_{i,p} : \| s_{i,p} - s_{j,n} \| < \epsilon\} \tag{2.63}$$

式中，$|u_{\epsilon}(s_{j,n})|$ 为邻域中点的个数，对具有空邻域的孤立点，取 s_i^l 的样本均值作为 $x_{(j-1)l+n+1}^{*}$ 的估计。

因此，给定 s_i^l，序列 s_j^l 的均方根预测误差 $\gamma(s_i^l, s_j^l)$ 定义为

$$\gamma(s_i^l, s_j^l) = \sqrt{\frac{1}{l-m} \sum_{n=m}^{l-1} (x_{(j-1)l+n+1}^{*} - x_{(j-1)l+n+1})^2} \tag{2.64}$$

当 $s_i^l = s_j^l$ 时，就是通常的样本外预测误差。对平稳时间序列，预期 $\gamma(s_i^l, s_j^l)$ 关于 i 和 j 是独立的，对角项 $\gamma(s_i^l, s_i^l)$ 将小于非对角项 $\gamma(s_i^l, s_j^l)$。一般情况下，$\gamma(s_i^l, s_j^l) \neq \gamma(s_j^l, s_i^l)$。

2.5 基于观测时间序列的系统非线性性检验

由于混沌现象必然发生在非线性系统内，要用混沌理论和方法从观测或实验获得时间序列中去分析原复杂系统，必须检验观测或实验获得的时间序列是否来自于非线性系统。因此，在对观测时间序列进行相空间重构之前，首先面临的一个问题是：获得的时间序列是源自于线性系统还是非线性系统？

假设实际问题中的复杂系统由式(2.9)的差分方程或式(2.10)的微分方程描述，系统状态不可直接观测，但可通过式(2.11)或式(2.12)进行观测。当式(2.9)或式(2.10)中的 F 是线性时，称动力系统是线性的；当 F 是非线性时，称动力系统是非线性的。当函数 F 和 h 至少有一是非线性时，时间序列 $\{x_n\}$ 是非线性的，这意味着时间序列的非线性性不等价于动力系统的非线性性，因为时间序列的非线性性可能是由于观测函数 h 而不是系统本身的函数 F 造成的。

基于观测时间序列的系统非线性性检验的目标是从观测或实验获得的时间序列 $\{x_n\}$ 中，尽可能多地去获取函数 F 的信息。需要解决的问题是 F 是线性的还是非线性的。当且仅当 F 是非线性时，才可预期动力系统具有感兴趣的确定性成分，如非线性极限环或混沌吸引子(确定性混沌永远不会出现在线性系统中)。

检验观测或实验获得的时间序列是否来自于非线性系统的一种方法是 J. Theiler 提出的替代数据法[62,63]。这种方法的基本思想是：对某一观测时间序列，限定某些线性过程作为零假设，零假设定义了候选的一种过程，它也许可能也许不可能恰当地解释原时间序列，然后根据零假设生成相应的替代时间序列，分别计算原时间序列和根据零假设生成的多组替代时间序列的判别统计量

(discriminating statistic)或称检验统计量(test statistic)。如果原时间序列的判别统计量不在零假设对应的多组替代时间序列的判别统计量的分布中，则可拒绝零假设，由此可检验动力系统的非线性性。

替代数据法检验观测时间序列是否来自于非线性系统的过程可以用图 2.7 表示。

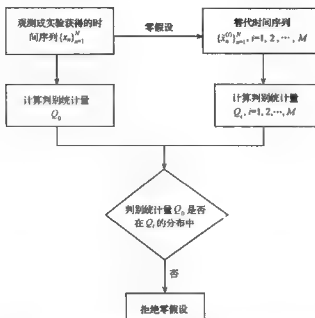


图 2.7　替代数据法检验时间序列是否来自于非线性系统的流程

由图 2.7 可以发现，利用替代数据法检验观测时间序列是否来自于非线性系统需要解决以下三个问题：

1）如何根据零假设生成替代时间序列？

2）选取什么量作为判别统计量？

3）如何判断是否拒绝零假设？

下面具体介绍这三个问题是如何解决的。

2.5.1　零假设及替代时间序列的约束生成算法

零假设是对时间序列特征的一种假设性猜测，它是产生替代时间序列的依据。不同的零假设对应不同的替代时间序列生成算法，其基本思想就是把原时间序列当成满足零假设的某一随机过程的一个实现，替代时间序列生成算法就是研究如何生成该过程的其他实现。通常的有两种实现方式，即经典实现(typical

realization)和约束实现(constrained realization)。经典实现就是利用传统的自助法提取详细的模型方程,然后运行产生 Monte Carlo 样本,只要模型方程能成功地提取,这种方法可以非常成功地计算置信区间。但约束实现方法更适合假设检验,它通过对原时间序列直接进行随机化,避免了拟合模型方程,建立在 Fourier 变换基础上的替代时间序列生成法就是一种约束实现方法。由约束实现生成的替代时间序列是一种随机时间序列,但要求它与原时间序列有相同的样本自相关函数(因此也有相同的功率谱)和相同的幅度分布,即满足下面两个条件:

1) $R_x(\tau)=R_{\tilde{x}(\tau)}, \tau=1,2,\cdots,\tau_{\max}$,其中 R 是样本自相关函数,$\tau_{\max}$是充分大的延迟时间间隔。

2) $A_x(x)=A_{\tilde{x}}(\tilde{x})$(或 $F_x(x)=F_{\tilde{x}}(\tilde{x})$),其中 $A_x(x)$是观测时间序列$\{x_n\}_{n=1}^N$的先验幅度分布($F_x(x)$是$\{x_n\}_{n=1}^N$的样本累积密度函数)。

下面介绍不同零假设下用约束实现生成替代时间序列的几种不同方法。

1. Fourier 变换算法(FT)

零假设:原时间序列来自具有常系数的高斯线性随机过程,即

$$s_n=\sum_{i=1}^{p}a_i s_{n-i}+\sum_{j=0}^{q}b_j\eta_{n-j} \tag{2.65}$$

其中$\{\eta_n\}$是均值为零、方差为 σ^2 的高斯白噪声,即$\{\eta_n\}\sim N(0,\sigma^2)$,表明$\{s_n\}$是一个高斯白噪声激励的 ARMA$(p,q)$模型的输出序列。

设$\{x_n\}_{n=1}^N$为实测时间序列,做以下离散 Fourier 变换

$$y_k=\sum_{n=1}^{N}x_n\exp\left[-i\frac{2\pi(k-1)(n-1)}{N}\right],\ k=1,2,\cdots,N \tag{2.66}$$

则替代时间序列$\{\tilde{x}_n\}_{n=1}^N$为[62,63]

$$\tilde{x}_n=\frac{1}{N}\sum_{k=1}^{N}y_k\exp(i\alpha_k)\exp\left[i\frac{2\pi(k-1)(n-1)}{N}\right],\ n=1,2,\cdots,N \tag{2.67}$$

其中 $0\leqslant\alpha_k\leqslant2\pi, k=1,2,\cdots,N$ 是随机数。

由于实测时间序列$\{x_n\}_{n=1}^N$一般为实数,而由式(2.67)产生的替代时间序列$\{\tilde{x}_n\}_{n=1}^N$可能是复数,这样会造成原始数据中某些固有特性流失到虚部中,为了保证替代时间序列$\{\tilde{x}_n\}_{n=1}^N$也是实数,我们要求 $0\leqslant\alpha_k\leqslant2\pi, k=1,2,\cdots,N$ 满足以下关系(这里假设 N 是偶数,当 N 是奇数时可类似处理):

$$\alpha_1=\alpha_{\frac{N}{2}+1}=0 \tag{2.68}$$

$$\alpha_k=-\alpha_{N+2-k},\ k=2,3,\cdots,\frac{N}{2} \tag{2.69}$$

上面的关系可由图 2.8 形象地说明。

事实上,在式(2.68)、式(2.69)的条件下,对 $n=1,2,\cdots,N$ 有

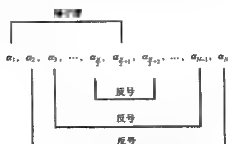


图 2.8　随机数 α_k 满足的关系

$$
\begin{aligned}
\widetilde{x}_n &= \frac{1}{N}\sum_{k=1}^{N}\left\{\sum_{m=1}^{N}x_m\exp\left[-\mathrm{i}\,\frac{2\pi(k-1)(m-1)}{N}\right]\right\} \\
&\quad \exp(\mathrm{i}\alpha_k)\exp\left[\mathrm{i}\,\frac{2\pi(k-1)(n-1)}{N}\right] \\
&= \frac{1}{N}\sum_{k=1}^{\frac{N}{2}}\exp(\mathrm{i}\alpha_k)\left\{\sum_{m=1}^{N}x_m\exp\left[\mathrm{i}\,\frac{2\pi(k-1)(n-m)}{N}\right]\right\} \\
&\quad + \frac{1}{N}\sum_{k=\frac{N}{2}+1}^{N}\exp(\mathrm{i}\alpha_k)\left\{\sum_{m=1}^{N}x_m\exp\left[\mathrm{i}\,\frac{2\pi(k-1)(n-m)}{N}\right]\right\} \\
&= \frac{1}{N}\sum_{k=1}^{\frac{N}{2}}\exp(\mathrm{i}\alpha_k)\left\{\sum_{m=1}^{N}x_m\exp\left[\mathrm{i}\,\frac{2\pi(k-1)(n-m)}{N}\right]\right\} \\
&\quad + \frac{1}{N}\sum_{j=2}^{\frac{N}{2}+1}\exp(\mathrm{i}\alpha_{N+2-j})\left\{\sum_{m=1}^{N}x_m\exp\left[-\mathrm{i}\,\frac{2\pi(j-1)(n-m)}{N}\right]\right\} \\
&= \frac{1}{N}\exp(\mathrm{i}\alpha_1)\sum_{m=1}^{N}x_m + \frac{1}{N}\sum_{k=2}^{\frac{N}{2}}\left\{\sum_{m=1}^{N}x_m\left[\exp\left(\mathrm{i}\left(\alpha_k + \frac{2\pi(k-1)(n-m)}{N}\right)\right)\right.\right. \\
&\quad \left.\left. + \exp\left(\mathrm{i}\left(\alpha_{N+2-k} - \frac{2\pi(k-1)(n-m)}{M}\right)\right)\right]\right\} + \frac{1}{N}\exp\left(\mathrm{i}\alpha_{\frac{N}{2}+1}\right)\sum_{m=1}^{N}x_m(-1)^{n-m} \\
&= \frac{1}{N}\sum_{m=1}^{N}x_m + \frac{2}{N}\sum_{k=2}^{\frac{N}{2}}\left\{\sum_{m=1}^{N}x_m\cos\left[\alpha_k + \frac{2\pi(k-1)(n-m)}{N}\right]\right\} + \frac{1}{N}\sum_{m=1}^{N}x_m(-1)^{n-m} \\
&= \text{实数}
\end{aligned}
\tag{2.70}
$$

其中第三个等式中用到了变换 $j-1=N-(k-1)$，即 $j=N+2-k$；第五个等式中用到了式(2.68)和式(2.69)。

关于随机数 α_k 的取法，文献[64]中对服从均匀分布的随机数(用 Monte Carlo 随机取余方法产生)和服从正态分布的随机数进行了数值模拟，表明服从不同分布

的随机数生成的替代时间序列对原时间序列的非线性性判别几乎没有影响，只要保证随机数的生成是充分随机的即可，因此，为简单起见，可取满足式(2.68)、式(2.69)的均匀分布的随机数 a_k。

利用 Fourier 变换的性质可知，由式(2.67)生成的替代时间序列$\{\tilde{x}_n\}_{n=1}^N$与原时间序列$\{x_n\}_{n=1}^N$有相同的自相关函数(等价地，有相同的 Fourier 功率谱)，但它们的分布幅度不一定相同。另外，拒绝这样的零假设，只能说明原时间序列是非高斯线性随机过程，即使具有非线性性，这种非线性性是由系统造成的，还是由观测造成的并不清楚，为此文献[63]提出了下面的幅度调节 Fourier 变换算法。

2. 幅度调节 Fourier 算法(AAFT)

零假设：原时间序列来自具有常系数的高斯线性随机过程经过单调静态非线性变换，即 $x_n=x(s_n)$，其中$\{s_n\}_{n=1}^N$是具有常系数的高斯线性随机过程，$x(\cdot)$称为"静态的"(static)或"即时的"(instantuous)，因而 x_n 只依赖于 s_n 而与 s_n 的过去值无关。另外还要求 $x(\cdot)$是单调函数，即 x^{-1}存在。

这样的零假设显然包含了 FT 算法中的零假设，且可用于检验系统的动态非线性性。

AAFT 算法的基本思想是：首先按比例排列原时间序列，使它与一个高斯分布一致，然后对重排的时间序列进行相位随机化，最后对相位随机化后的时间序列按比例重排，使它有原时间序列的幅度分布，具体步骤如下[63]。

步骤 1：模拟 x^{-1}(即记录与$\{x_n\}_{n=1}^N$有相同排列结构的白噪声，记为$\{r_n\}_{n=1}^N$)。首先生成一个高斯时间序列$\{g_n\}_{n=1}^N$，其中每一个元独立地从一个高斯伪随机数生成。记 rank(x_n)为 x_n 的升序阶数[例如，若 x_n 是$\{x_n\}_{n=1}^N$中的第 3 个最小，则 rank$(x_n)=3$]，则重新按比例排列的序列为

$$r_n = g(x_n) = g_{\mathrm{rank}(x_n)},\ n = 1,2,\cdots,N \tag{2.71}$$

步骤 2：对$\{r_n\}_{n=1}^N$做 Fourier 变换，并加一随机相位，然后做 Fourier 反变换，产生时间序列$\{\tilde{r}_n\}_{n=1}^N$。

步骤 3：模拟 x(即记录$\{\tilde{x}_n\}_{n=1}^N$使它与$\{\tilde{r}_n\}_{n=1}^N$有相同的排列结构)。根据原时间序列$\{x_n\}_{n=1}^N$的分布，由$\{\tilde{r}_n\}_{n=1}^N$的排列顺序，对 g 求逆，即

$$\tilde{x}_n = \tilde{g}(\tilde{r}_n) = x_{\mathrm{rank}(\tilde{r}_n)},\ n = 1,2,\cdots,N \tag{2.72}$$

这样得到的时间序列$\{\tilde{x}_n\}_{n=1}^N$为原时间序列$\{x_n\}_{n=1}^N$的替代时间序列。显然$\{\tilde{x}_n\}_{n=1}^N$与$\{x_n\}_{n=1}^N$的幅度分布是完全一致的，但它们的自关联函数只有在 $x(\cdot)$单调且步骤 1 和步骤 3 分别成功地模拟了 $x^{-1}(\cdot)$和 $x(\cdot)$的情况下才是相同的。

若 F_0 表示高斯累积密度函数($\{s_n\}_{n=1}^N$、$\{r_n\}_{n=1}^N$和$\{\tilde{r}_n\}_{n=1}^N$是相同的)，则由步骤 1

的记录可得 $r_n = g(x_n) = F_0^{-1}(F_x(x_n))$。由于 $x(\cdot)$的逆唯一，在标准化因子下(即如果 N 无限，$\{s_n\}_{n=1}^N$和$\{r_n\}_{n=1}^N$在标准化情况下是相等的)，g 和 x^{-1}一致，这样获得的$\{r_n\}_{n=1}^N$满足 $R_r(\tau)=R_s(\tau)-R_r(\tau)$，因此$\{s_n\}_{n=1}^N$，$\{r_n\}_{n=1}^N$和$\{\bar{r}_n\}_{n=1}^N$是同一高斯过程的实现，而且通过一个依赖于 g 的函数 φ 建立了 $R_s(\tau)$和 $R_x(\tau)$之间的关系，即 $R_x(\tau)=\varphi(R_s(\tau))$。相应地，由步骤 3 的记录可得 $\bar{x}_n = g^{-1}(\bar{r}_n) = F_x^{-1}(F_0(\bar{r}_n))$，且 $x = g^{-1}$，$R_{\bar{x}}(\tau)=\varphi(R_{\bar{r}}(\tau))=\varphi(R_s(\tau))$，这样 $R_{\bar{x}}(\tau)=R_x(\tau)$。

在实际应用中，一般 $x(\cdot)$不一定单调，即使 $x(\cdot)$单调，步骤 1 中的模拟也不可能完全精确，即 g 一般不是非线性测量函数 $x(\cdot)$的真正逆，而且式(2.72)中的 $\bar{g}$ 与 g^{-1}一般也不一样，因此这样生成的替代时间序列与原时间序列列自关联不可能完全一样，这样当判别统计量对自关联敏感时，自关联的偏差会导致错误地拒绝零假设。文献[65]指出，对具长相干时间(long coherent time)的时间序列，用 AAFT 算法会错误地拒绝零假设。

为了去掉对 $x(\cdot)$单调这一要求，文献[66]提出了下面的迭代算法。

3. 迭代的幅度调节 Fourier 算法(IAAFT)

零假设：原时间序列来自具常系数的高斯线性随机过程经过静态非线性变换，这里对非线性转换函数 $x(\cdot)$不要求可逆。

由 AAFT 算法，生成的替代时间序列与原时间序列的自相关函数有偏差，可以通过多次迭代的方法使这种偏差限定在预期的范围内。

设原时间序列为$\{x_n\}_{n=1}^N$，记

$$|y_k| = \left|\frac{1}{\sqrt{N}}\sum_{n=1}^{N} x_n \exp\left[\mathrm{i}\,\frac{2\pi(k-1)(n-1)}{N}\right]\right|,\ k=1,2,\cdots,N \quad (2.73)$$

又设$\{c_n\}_{n=1}^N$为原时间序列$\{x_n\}_{n=1}^N$按数值上升顺序的重排，则替代时间序列的生成算法过程如下。

步骤 1：由 AAFT 算法或对原时间序列的一个简单随机重排产生初始替代时间序列$\{\bar{x}_n^{(0)}\}_{n=1}^N$。

步骤 2：迭代。分为两个连续的过程：

1) $\bar{x}_n^{(i)} \to r_n^{(i)}$ 是 Fourier 域中的粗糙的滤波。即对$\{\bar{x}_n^{(i)}\}_{n=1}^N$做离散 Fourier 变换

$$\tilde{y}_k^{(i)} = \frac{1}{\sqrt{N}}\sum_{n=1}^{N} \bar{x}_n^{(i)} \exp\left[\mathrm{i}\,\frac{2\pi(k-1)(n-1)}{N}\right],\ k=1,2,\cdots,N \quad (2.74)$$

然后做反变换，但 y_k 用代替 $\tilde{y}_k^{(i)}$，且保持相位 $\exp(\mathrm{i}\Psi_k^{(i)})=\tilde{y}_k^{(i)}/|\tilde{y}_k^{(i)}|$，即

$$r_n^{(i)} = \frac{1}{\sqrt{N}}\sum_{k=1}^{N} |y_k| \exp(\mathrm{i}\Psi_k^{(i)}) \exp\left[-\mathrm{i}\,\frac{2\pi(k-1)(n-1)}{N}\right],\ n=1,2,\cdots,N \quad (2.75)$$

2) $r_n^{(i)} \to \tilde{x}_n^{(i+1)}$ 由重新排列顺序产生，即

$$\tilde{x}_n^{(i+1)} = c_{\text{rank}(r_n^{(i)})}, \quad n = 1,2,\cdots,N \qquad (2.76)$$

由迭代过程可知，在迭代的第一步中，$\tilde{x}_n^{(i)}$ 与 $r_n^{(i)}$ 的功率谱相同(相应地自关联也相同)，在迭代的第二步中，$r_n^{(i)}$ 与 $\tilde{x}_n^{(i+1)}$ 的幅度分布一致。但第一步迭代中保持的功率谱在第二步迭代中被改变，因此需多次迭代，在每次迭代后，检查功率谱的剩余偏差，直至达到要求的精度为止。最后能达到的精度依赖于原始数据的大小和结构，一般多次迭代后足以保证假设测试的可信度。

如果通过 i 次迭代以后，以 $\{\tilde{x}_n^{(i+1)}\}_{n=1}^N$ 作为 $\{x_n\}_{n=1}^N$ 的替代时间序列，则替代时间序列与原时间序列有相同的幅度分布，但自关联可能发生偏差，文献[67]记这种迭代算法为 IAAFT-1；如果要求自关联一致，而允许幅度分布有偏差，可以以 $\{r_n^{(i)}\}_{n=1}^N$ 作为 $\{x_n\}_{n=1}^N$ 的替代时间序列，文献[67]记这种算法为 IAAFT-2。文献[67]证实，无论 IAAFT-1 算法还是 IAAFT-2 算法获得的替代时间序列要比用 AAFT 算法获得的替代时间序列与原时间序列的自关联偏差小得多。关于 IAAFT 算法的更详细的讨论可见文献[67]。

4. 改进的幅度调节 Fourier 算法(CAAFT)

为了去掉 AAFT 算法的零假设中对 $x(\cdot)$ 单调这一要求，文献[68]从另外角度提出了下面的改进的幅度调节 Fourier 算法。

零假设：原时间序列来自具常系数的高斯线性随机过程经过静态非线性变换，但这一变换不一定是单调的。

CAAFT 算法的过程如下[68]：

步骤 1：对观测时间序列 $\{x_n\}_{n=1}^N$，由 AAFT 算法，生成 $\{r_n\}_{n=1}^N$、$\{\tilde{r}_n\}_{n=1}^N$ 和 $\{\tilde{x}_n\}_{n=1}^N$。

步骤 2：选择 $\tau_{\max}$，对 $\tau=1,2,\cdots,\tau_{\max}$，分别计算 $R_{\tilde{x}}(\tau)$、$R_x(\tau)$ 和 $R_r(\tau)$。

步骤 3：在图 $R_r(\tau)-R_{\tilde{x}}(\tau)$ 中，通过线性插值，获得分段线性函数 $\hat{\Psi}$，使 $R_r(\tau)=\hat{\Psi}(R_{\tilde{x}}(\tau))$。

步骤 4：由 $R_u(\tau)=\hat{\Psi}(R_x(\tau))$ 计算 $R_u(\tau)$。

步骤 5：选择 AR 模型的一个合适的阶 p(如 $p=\tau_{\max}$)，并从 $R_u(\tau)$ 中估计 AR(p)的系数。

步骤 6：利用稳定的 AR 模型生成 $\{u_n\}_{n=1}^N$。

步骤 7：把 $\{u_n\}_{n=1}^N$ 转换成 $\{w_n\}_{n=1}^N$，即 $w_n=F_x^{-1}(F_\Phi(u_n))=s_{\text{rank}(u_n)}$。

步骤 8：步骤 1 至步骤 8 重复 K 次，产生 K 组时间序列 $\{w_n^{(1)}\}_{n=1}^N,\{w_n^{(2)}\}_{n=1}^N,\cdots,\{w_n^{(K)}\}_{n=1}^N$。

步骤 9：分别计算$\{w_n^{(1)}\}_{n=1}^N,\{w_n^{(2)}\}_{n=1}^N,\cdots,\{w_n^{(K)}\}_{n=1}^N$的自关联$R_w^{(1)},R_w^{(2)},\cdots,R_w^{(K)}$，找出最接近$R_x$的$R_w^{(k)}$。

步骤 10：对应选择的第k次试验，利用 AR 模型，步骤 6 和步骤 7 重复M次，生成M组替代时间序列。

由 CAAFT 算法生成的替代时间序列与原时间序列有相同的幅度分布，且有近似相同的自关联，因此克服了 AAFT 算法生成的替代时间序列与原时间序列的自关联偏差这一缺陷，用 CAAFT 算法生成的替代时间序列用于检验原时间序列是否来自于非线性系统时，要比用 AAFT 算法生成的替代时间序列更精确。由 CAAFT 算法生成的替代时间序列与原时间序列的自关联的近似是无偏的，这种近似的方差与 AAFT 算法产生的方差一样大，但大于 IAAFT 算法产生的方差，因此用 CAAFT 算法比 IAAFT 算法更稳健，意味着如果用 CAAFT 算法生成的替代时间序列通过了原时间序列的非线性性检验，则这样的非线性性证据比用 AAFT 算法或 IAAFT 算法获得的结果更可靠。

5．一般的约束随机化方法(GCR)

在前面的各种替代时间序列的生成算法中，零假设通常根据原时间序列的一个完全可观测集的性质限定。例如，零假设为原时间序列来自具有常系数的高斯线性随机过程，包括一个可能可逆也可能不可逆的静态非线性变换，所以拒绝零假设并不能说明原时间序列一定来自于非线性动力系统，它可能来自非高斯非线性随机过程或来自非静态测量函数(如$x_n=s_n^2s_{n-1}$)或来自线性随机非平稳过程等。文献[69，70]的研究表明，对非最小相位的线性非高斯过程，用 IAAFT 算法生成的替代时间序列会导致错误地拒绝零假设，文献[65]的研究也表明，对循环平稳(cyclostationary)的线性高斯过程，用 FT 算法或 AAFT 算法生成的替代时间序列也会导致错误地拒绝零假设。为了解决这一问题，文献[71]提出了一般的约束随机化方法，它可以检验原时间序列的任何完全可观测集的性质导致的零假设，包括多变量时间序列、非平稳性、非线性性或其他约束。其基本思想是：首先根据某个成本函数$E(\{\tilde{x}_n\})$，指定约束$F_i(\{\tilde{x}_n\})=0$，当约束满足时，构造的成本函数有一个全局最小解，然后利用模拟退火(simulated annealing)法，对时间序列$\{x_n\}_{n=1}^N$的所有排列$\{\tilde{x}_n\}_{n=1}^N$，求$E(\{\tilde{x}_n\})$的最小解。

设$\{\tilde{x}_n\}_{n=1}^N$表示$\{x_n\}_{n=1}^N$的替代时间序列，约束的通常形式为

$$F_i(\{\tilde{x}_n\})=0,\ i=1,2,\cdots,I \tag{2.77}$$

这种约束可转换成一个成本函数

$$E(\{\tilde{x}_n\})=\Big(\sum_{i=1}^{I}|\omega_iF_i(\{\tilde{x}_n\})|^q\Big)^{\frac{1}{q}} \tag{2.78}$$

当约束满足时，$E(\{\tilde{x}_n\})$有全局最小这一事实不受权重 $\omega_i\neq0$ 及阶 q 的影响。在 $q=2$时，获得最小二乘或 L^2 平均；在 $q=1$ 时，获得 L^1 平均；在 $q\to\infty$时，获得最大距离；而当 q 趋向于零时，几何平均也是可能的。

在具体实现时，例如可考虑替代时间序列$\{\tilde{x}_n\}_{n=1}^N$的自关联函数 $\tilde{C}(\tau)=\langle\tilde{x}_n\tilde{x}_{n-\tau}\rangle$与原时间序列$\{x_n\}_{n=1}^N$的自关联函数 $C(\tau)=\langle x_nx_{n-\tau}\rangle$相同这一要求，则可限定约束为

$$F_\tau(\{\tilde{x}_n\})=\tilde{C}(\tau)-C(\tau),\ \tau=1,2,\cdots,\tau_{\max} \tag{2.79}$$

这样成本函数可以是

$$E(\{\tilde{x}_n\})=\max\{|\tilde{C}(\tau)-C(\tau)|:\tau=1,2,\cdots,\tau_{\max}\} \tag{2.80}$$

或

$$E(\{\tilde{x}_n\})=\left(\sum_{\tau=1}^{\tau_{\max}}|\tilde{C}(\tau)-C(\tau)|^q\right)^{\frac{1}{q}} \tag{2.81}$$

在这一方法中，替代时间序列的选取是在原时间序列中取相同的值但不同的时间阶，这可保证替代时间序列和原时间序列取自同一(未知)密度分布的随机过程，通过对原时间序列重新排序可完成这一步。为了对$\{x_n\}_{n=1}^N$的各种不同排序求成本函数的最小解，可以利用模拟退火法，通过交换$\{x_n\}_{n=1}^N$中的次序，构形被更新，退火计划将确定是接受还是拒绝，在一个合适的冷却计划下，退火过程可以达到任何要求的精度。当然除了模拟退火法外，遗传算法(genetic algorithms)也可用于解决这类问题。

对中等长度和较长的时间序列，对给定的约束，GCR 算法生成的替代时间序列可以非常精确，但这是以非常大的计算时间为代价的，因此这种方法在实际应用时仍受到限制。

2.5.2 判别统计量的选取

判别统计量是定量表征时间序列某个特性的特征量，它是判别实测时间序列与零假设是否符合的依据。如果实测时间序列的判别统计量与由零假设所生成的替代时间序列的判别统计量不相符，则认为零假设不成立，从而说明实测时间序列具有零假设以外的特征。如果替代时间序列是用前面的约束方法生成的，理论上判别统计量可以是任何非线性测度。

常用的判别统计量有以下非线性量。

(1) 最大 Lyapunov 指数

从观测或实验获得的时间序列计算最大 Lyapunov 指数的方法见 2.3.3 节。

(2) 关联维数

从观测或实验获得的时间序列计算关联维数的方法见 2.3.1 节。但用于判别

统计量时更多的是用关联维数的极大似然估计，也称 Takens 估计量[72]，即

$$D_{\text{takens}}(m,\tau,r)=\frac{C_N(r,\tau,m)}{\int_0^r \frac{C_N(\varepsilon,\tau,m)}{\varepsilon}\mathrm{d}\varepsilon} \tag{2.82}$$

式中，$C_N(r,\tau,m)$是 2.3.1 节定义的关联积分。

(3) 非线性预测误差

设观测或实验获得的时间序列为$\{x_n\}_{n=1}^N$，做式(2.14)的相空间重构，记 $\boldsymbol{x}_{n_k}$，$k=1,2,\cdots,K$ 为 $\boldsymbol{x}_n$ 的 K 个最近邻点，则 x_{n+1}的一步预测为

$$\hat{x}_{n+1}=\sum_{k=1}^{K}\omega_k x_{n_k+1} \tag{2.83}$$

其中

$$\omega_k=\frac{\|\boldsymbol{x}_n-\boldsymbol{x}_{n_k}\|^{-2}}{\sum_{i=1}^{K}\|\boldsymbol{x}_n-\boldsymbol{x}_{n_i}\|^{-2}} \tag{2.84}$$

因此平方平均意义下的一步预测误差为

$$\gamma(m,\tau,K)=\sqrt{\frac{1}{N-2}\sum_{n=0}^{N-2}(x_{n+1}-\hat{x}_{n+1})^2} \tag{2.85}$$

可作为判别统计量，它具有计算简便的优点。

2.5.3　统计检验方法

如果能获得观测时间序列的多次实现，则直接利用 Kolmogorov-Smirnov 或 Mann-Whitney U 检验比较实测时间序列和替代时间序列的两个完全分布或利用 Student t 检验只比较实测时间序列和替代时间序列的分布均值都是可能的。通过把实测时间序列分成几个短的时间序列也可获得实测时间序列的多次实现，但如果实测时间序列本身就不长，像关联维数、Lyapunov 指数等判别统计量的计算会受到影响。实际问题中的实测时间序列一般只能实现一次，因此用下面的两种统计检验方法。

1. Sigma 检验

计算实测时间序列的某种判别统计量 Q_0 和多组替代时间序列相应的这种判别统计量的均值 $\overline{Q}_s$ 和标准差 σ_s，定义显著性度量为

$$\mathrm{DIFF}(Q)=\frac{|Q_0-\overline{Q}_s|}{\sigma_s} \tag{2.86}$$

假定满足零假设的时间序列的判别统计量服从高斯(正态)分布，则可以通过正态分布的参数检验拒绝或接受零假设，其显著性水平 p 值为

$$p = \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right) = 2[1 - \Phi(x)] \tag{2.87}$$

其中

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \tag{2.88}$$

查表可得，$p=0.05$ 时，$x=1.96$，表明若取显著性水平 $p=0.05$，则当 $\mathrm{DIFF}(Q) \geqslant 1.96$ 时，观测时间序列与替代时间序列有明显差别，观测时间序列以 95%的置信水平拒绝零假设。当替代时间序列的组数越多时，获得的结论越可信。

2. Rank-Order 检验

由于满足零假设的时间序列的判别统计量不一定服从正态分布（例如，零假设是某个线性高斯过程的单调静态非线性变换时），因此用 Sigma 检验这种参数估计方法可能会产生误判，一种更可靠的方法是下面的 Rank-Order 检验[63,73]。

假定零假设是真的时拒绝零假设的概率为 α，即对应的显著性水平为 $(1-\alpha)\times 100\%$，则对单边测试，生成 $M=\frac{k}{\alpha}-1$ 组替代时间序列，其中 k 是正整数，这样，包括原时间序列，有 $\frac{k}{\alpha}$ 组时间序列，则每一组的判别统计量都等于全部判别统计量的最小值的概率为 α，即如果观测时间序列的判别统计量为 $\frac{k}{\alpha}$ 组时间序列中的最小，则拒绝零假设。对双边测试，生成 $M=\frac{2k}{\alpha}-1$ 组替代时间序列，则每一组的判别统计量都等于全部判别统计量的最小值或最大值的概率为 α，即如果实测时间序列的判别统计量为 $\frac{2k}{\alpha}$ 组时间序列中的最小或最大，则拒绝零假设。当 k 的值较大时，测试结果比 $k=1$ 更好。为了减少计算量，一般取 $k=1$，这样为了 95%的显著性水平，分别需要 19 或 39 组替代时间序列用于单边或双边测试。

2.6　基于观测时间序列的系统确定性检验

如果系统的将来值可由过去值完全确定，则称系统是确定性的。例如，式(2.9)这样的差分方程或式(2.10)这样的微分方程描述的系统，只要函数 F 是单值的，则这样的系统的将来状态 z_{n+1} 或 $z(t+\Delta t)$ 由过去状态 z_n 或 $z(t)$ 完全确定，因而是确定性的。但对只能获得观测时间序列的动力系统，这样的 F 是未知的，因此要从时间序列本身来判断它是由确定性机制还是随机性机制所产生的。不同机制产生的时间序列在建立模型时所用的方法是不同的，如果是由随机性机制所产生

的,就应该采用统计学的方法进行研究,如果是由确定性机制产生的低维混沌系统,那么,无论多么高阶次的线性模型都不可能很好地解释该系统,相反一个低阶非线性模型或许能很好地刻画原动力系统。

2.6.1 从非线性预测判断系统的确定性

对确定性系统,由于将来的状态可由以前的状态完全确定,因此理论上讲是完全可以预测的;但当确定性系统具有混沌特征时,由于对初值的敏感依赖性,测量误差会不断放大,因此虽然也可预测,但预测并不精确;随机系统是完全无法预测的。因此对观测时间序列,可以用下面的思想来判断它是否来自于确定性系统:利用观测数据构造动力系统模型,然后考察据此模型进行的预测是否精确。如果预测是完美的,则系统是完全确定的;如果预测较好,但并不完美,则系统具有确定性成分;如果预测非常糟糕,则系统没有确定性。

从观测时间序列中构造动力系统模型的方法有很多,从预测的角度看,一种简单的方法基于以下思想:假设 $z_{n+1}=F(z_n)$ 中的 F 是连续的,则对现在状态 z_n,寻找与其最接近的状态 $z_{\eta(n)}$,由于 F 的连续性,保证了 z_{n+1} 与 $z_{\eta(n)+1}$ 也非常接近。

假设通过某个未知的观测函数 $h(\cdot)$ 可以观测到单变量时间序列 $\{x_n\}_{n=1}^N$,作式(2.14)的相空间重构。为了预测时间序列 $\{x_n\}_{n=1}^N$ 的下一个值 x_{N+1},取 N 时刻的嵌入向量 $x_N=(x_N,x_{N-\tau},\cdots,x_{N-(m-1)\tau})$,在 $x_{N_0},x_{N_0+1},\cdots,x_{N-1}$ 中寻找最靠近 x_N 的向量(为了计算简便,可取最大范数作为距离定义),不妨记为 $x_{\eta(N)}$,即

$$\| x_n - x_{\eta(N)} \| = \min\{ \| x_n - x_j \| : j = N_0, N_0+1, \cdots, N-1\} \qquad (2.89)$$

如果 x_N 与 $x_{\eta(N)}$ 充分靠近,则 x_{N+1} 与 $x_{\eta(N)+1}$ 也会非常接近,因此 x_{N+1} 的预测值 $\hat{x}_{N+1}$ 可取为 $x_{\eta(N)+1}$,即

$$\hat{x}_{n+1} = x_{\eta(N)+1} \qquad (2.90)$$

这一简单预测模型有许多改进,例如取定 x_N 后,不是找一个最靠近它的向量 $x_{\eta(N)}$,而是找与 x_N 最接近的 K 个向量 x_{N_k}, $k=1,2,\cdots,K$,取 x_{N+1} 的预测值 $\hat{x}_{N+1}$ 为 $x_{N_1+1},x_{N_2+1},\cdots,x_{N_K+1}$ 的平均值,即

$$\hat{x}_{n+1} = \frac{1}{K}\sum_{k=1}^{K} x_{N_k+1} \qquad (2.91)$$

或者在取定 x_N 后,对某个 ϵ,记 $N(\epsilon,x_N)$ 为以 x_N 为中心 ϵ 为半径的邻域,即若 $x_n\in N(\epsilon,x_N)$,则 $\| x_n - x_N \| < \epsilon$,取 x_{N+1} 的预测值 $\hat{x}_{N+1}$ 为

$$\hat{x}_{N+1} = \frac{1}{|N(\epsilon,x_n)|}\sum_{x_n\in N(\epsilon,x_n)} x_{n+1} \qquad (2.92)$$

其中 $|N(\epsilon,x_n)|$ 表示邻域 $N(\epsilon,x_n)$ 中向量 x_n 的个数。当邻域内没有向量与 x_N 的距离小于 ϵ 时,则增加 ϵ,直到找到某些向量为止。

为了检查预测值 $\hat{x}_{N+1}$ 的正确性，可以通过与观测值 x_{N+1} 进行比较，定义预测误差为 x_{N+1} 与 $\hat{x}_{N+1}$ 的差。一次预测可能较好或较差，偶然性较大，为了克服这种偶然性，可以取多次预测误差的平均。假设有长为 $N+L$ 的观测时间序列 $\{x_n\}_{n=1}^{N+L}$，取前 N 个数据为构造模型所需，利用此模型来预测后 L 个数据。记 $\hat{x}_{N+l}$ 为 x_{N+l} 的预测值，则可得平均预测误差

$$\varepsilon = \frac{1}{L}\sum_{l=1}^{L}(x_{N+l} - \hat{x}_{N+l})^2 \tag{2.93}$$

如果 ε 非常大，意味着预测不好，因此时间序列不是来自于确定性系统；反之，如果 ε 较小，说明时间序列可能来自于确定性系统。

在利用前 N 个数据预测 $x_{N+l}, l=1,2,\cdots,L$ 时，可以有以下两种不同的方法：

1）利用前面的方法预测 $N+1$ 时刻的值 $\hat{x}_{N+1}$，然后利用预测值 $\hat{x}_{N+1}$ 构造一个新的嵌入向量

$$\boldsymbol{x}_{N+1} = (\hat{x}_{N+1}, x_{N+1-\tau}, \cdots, x_{N+1-(m-1)\tau}) \tag{2.94}$$

找与 $\boldsymbol{x}_{N+1}$ 最近的向量，预测 $N+2$ 时刻的值 $\hat{x}_{N+2}$，然后再利用预测值 $\hat{x}_{N+2}$ 构造一个新的嵌入向量

$$\boldsymbol{x}_{N+2} = (\hat{x}_{N+2}, x_{N+2-\tau}, \cdots, x_{N+2-(m-1)\tau}) \tag{2.95}$$

找与 $\boldsymbol{x}_{N+2}$ 最近的向量，预测 $N+3$ 时刻的值 $\hat{x}_{N+3}$，这样一直下去。这种方法事实上可用于外推到对超过观测时间序列长度的时刻进行预测，虽然这种方法在用于预测不远的将来时刻的值时是有用的，但在判别时间序列是否来自于确定性系统时最好还是直接用已知的观测数据，即下面的方法。

2）为了预测 $N+2$ 时刻的值 $\hat{x}_{N+2}$，构造嵌入向量

$$\boldsymbol{x}_{N+1} = (x_{N+1}, x_{N+1-\tau}, \cdots, x_{N+1-(m-1)\tau}) \tag{2.96}$$

这时虽然 $N+1$ 时刻的预测值 $\hat{x}_{N+1}$ 已经得到，但还是用 x_{N+1}，因为它来自于实际观测，同理以 x_{N+2} 而不是 $\hat{x}_{N+2}$ 构造嵌入向量，依此下去。

最后要说明的是以什么标准衡量平均预测误差 ε 是大还是小，仅仅从 ε 的数值本身无法确定，因为 ε 与时间序列 $\{x_n\}_{n=1}^{N+L}$ 的数值大小有关，而不同问题中 $\{x_n\}_{n=1}^{N+L}$ 的数量级是不一样的。为了解决这一问题，考虑如果对每一时刻 $N+l$ 都取相同的预测值 $\hat{x}_{\text{same}}$，使平均预测误差

$$\varepsilon_{\text{same}} = \frac{1}{L}\sum_{l=1}^{L}(x_{N+l} - \hat{x}_{\text{same}})^2 \tag{2.97}$$

最小，则由最小二乘法可知

$$\hat{x}_{\text{same}} = \frac{1}{L}\sum_{l=1}^{L} x_{N+l} \tag{2.98}$$

即这样的预测值是后 L 个数据的均值，因此此时的平均预测误差 $\varepsilon_{\text{same}}$ 就是后 L 个数据的方差 σ^2。由于这种预测是比较差的，因此要判断式(2.93)的平均预测误差的

大小，可让它与后 L 个数据的方差 σ^2 进行比较，即若 ε/σ^2 接近于1，则表明平均预测误差较大，若 ε/σ^2 接近于0，则表明平均预测误差 ε 较小。

这种方法的相关内容及应用例子可见文献[74]。

2.6.2　利用递归图判断系统的确定性

理论上讲，任何时间序列不是周期的，就是非周期的，然而当出现观测噪声时，从周期系统中观测的时间序列可能出现非周期性。非周期性意味着状态变量永远不会严格地回到它们以前的值，但在一个非周期系统中，状态变量可能回到非常靠近以前某时刻的状态，因此可通过刻画这种靠近程度和刻画经过多久才会出现这种靠近等来描述非周期程度。

由于不能直接测量系统的所有状态变量，需要利用嵌入技术把获得的测量数据表示为状态变量。设 $\{x_n\}_{n=1}^{N}$ 为观测获得的时间序列，做式(2.14)的相空间重构，得向量 $\boldsymbol{x}_n$，每一个 $\boldsymbol{x}_n$ 是 m 维嵌入空间中的一个点，代表了系统在时刻 n 的状态。

计算 i 和 j 两个时刻 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 的距离 $\delta_{ij}=\|\boldsymbol{x}_i-\boldsymbol{x}_j\|$，如果时间序列以 T 为周期，则当 $|i-j|=nT(n=0,1,\cdots)$ 时，$\delta_{ij}=0$；相反，对非周期时间序列，δ_{ij} 就没有这样的性质。为此考虑以下问题：假设取定某个距离 r，什么时候会出现 $\|\boldsymbol{x}_i-\boldsymbol{x}_j\|<r$。解决这一问题的一种方法是以 i 作为横坐标，以 j 作为纵坐标作图，当 $\|\boldsymbol{x}_i-\boldsymbol{x}_j\|<r$ 时，在 (i,j) 处作一个点，这样的图叫做递归图(recurrence plots)，它描述了重构的轨线是如何递归或重复其自身的，揭示了系统的时间关联信息。当 r 固定时，递归图关于对角线对称，如果在不同的点处 r 取得不一样，递归图也可以不对称。

例2.3　二次映射

$$x_{n+1}=3.5x_n(1-x_n) \tag{2.99}$$

它的轨线是以4为周期的，当 $m=2,\tau=1,r=0.001$ 时，其递归图见图2.9，$m=2,\tau=1,r=0.01$ 时，其递归图见图2.10。

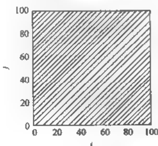


图2.9　式(2.99)在 $r=0.001$ 时的递归图

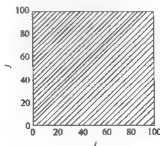


图2.10　式(2.99)在 $r=0.01$ 时的递归图

例 2.4　混沌系统

$$x_{n+1} = 4x_n(1 - x_n) \tag{2.100}$$

当 $m=2, \tau=1, r=0.001$ 时的递归图见图 2.11，而当 $m=2, \tau=1, r=0.01$ 时的递归图见图 2.12。

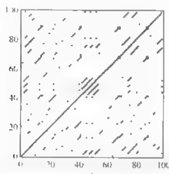


图 2.11　式(2.100)在 $r=0.001$ 时的递归图

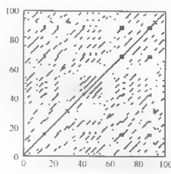


图 2.12　式(2.100)在 $r=0.01$ 时的递归图

例 2.5　随机白噪声当 $m=2, \tau=1, r=0.1$ 时的递归图见图 2.13，而当 $m=2, \tau=1, r=1$ 时的递归图见图 2.14。

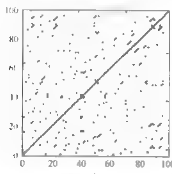


图 2.13　随机白噪声在 $r=0.1$ 时的递归图

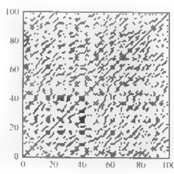


图 2.14　随机白噪声在 $r=1$ 时的递归图

由上述例子可知，对来自于周期系统的时间序列，递归图为与对角线平行的一些带构成，这些带之间的间距为周期 T。对来自于混沌系统的时间序列，递归图有更复杂的结构，有一些几乎周期的轨线的迹象，这可以从一些与对角线平行的、小的带中发现；来自于随机系统的时间序列就没有这种结构，其递归图几乎是毫无规律的。因此递归图也可以从某种程度上刻画系统的确定性特征。

递归图的思想最早是由 J. P. Eckmann 等在文献[75]中提出的，关于递归图的更多内容可见文献[76～82]。

2.7　观测时间序列噪声处理技术

观测和实验获得的时间序列通常包含噪声，噪声的存在一方面会淡化系统的确定性，即降低确定规律的显著性；另一方面又可能提供一些“虚假的规律或者确定性”，从而严重影响时间序列的预测效果。通过观测或实验获得的时间序列中主要含有测量噪声和系统噪声两种类型的噪声。

测量噪声(measurement noise)：指观测量被误差污染，它可能由测量仪器引起，与动力系统本身无关。设动力系统满足 $z_{n+1}=F(z_n)$，测量的量为 $x_n=h(z_n)+\omega_n$，其中 $h(\cdot)$是把吸引子中的点映为实数的光滑函数，ω_n 是随机数，序列$\{\omega_n\}$称为测量噪声。这类噪声是伴随着时间序列在观测的时候被叠加进来的，也称为外部噪声，一般是可以去掉的。

系统噪声(dynamical noise)：指系统内小的随机数在每一步的干扰所产生的反馈，它由主要系统参数小的随机波动的传播导致随机状的波动引起，也可能由系统内在事件随机发生引起，它直接影响系统的实时演化。动力系统满足 $z_{n+1}=F(z_n+\eta_n)$。这类噪声是与外部环境完全无关的，即在动力系统演化的过程中噪声作为一种输入进入系统，参与并影响系统的演化，这种噪声也称为内部噪声。内部噪声无法通过简单的滤波技术处理，需要借助系统的演化特征进行处理。

噪声对数据分析技术的影响的严重程度很大地依赖于噪声的级别和属性。例如，在数据处理过程中，系统噪声比测量噪声引起的问题更严重，这是由于有测量噪声的情况下，确定性系统中相邻的干净的轨线总存在。更进一步，系统噪声有时被解释为动力系统的具小幅度的高维确定性部分。即使不是这样，由于系统噪声会诱导或延迟动力系统通向不同的行为(例如分叉)，因此系统噪声对观测的动力系统可能是本质性的。

在对观测时间序列进行分析时，噪声的存在会产生很多不利的影响，这些影响包括：

1）吸引子的自相似性被破坏。

2）相空间重构后在小标度区间上出现高维。

3）相邻轨线的发散不是指数的。

4）不管使用哪种预测方法和记录多少位的数据，预测误差总有下界。

噪声对关联维数的影响要小于对 Lyapunov 指数、熵等的影响，尤其小于对预测的影响。噪声对关联维数、Lyapunov 指数、熵和预测的影响程度很大地依赖于噪声的级别和噪声的性质。一般地，当噪声级接近于几个百分点时，估计可能变得十分不可信。为此在对观测时间序列进行分析前，需要进行降噪处理，目前的降噪方法主要适用于测量噪声，还没有技术可除去系统噪声，实际上，系统噪声是否应该除去也是值得讨论的问题。

2.7.1　噪声级别的估计

通常,可以使用信噪比来描述信号和噪声的关系,信号强度和噪声强度的对比就叫"信号噪声比"(signal to noise ratio)或者简称 S/N。S/N 在习惯上也经常以 dB 为单位来表示。这里将使用噪声强度来对信号和噪声之间的强度对比关系进行描述。定义噪声级别为 z,假设含有测量噪声(假设为高斯噪声)的信号为 y,$y=x+\xi$,x 是原数据,ξ 代表噪声,则 $z=\sigma_\xi^2/\sigma_x^2$,$\sigma^2$ 是相应数据的方差。

估计噪声级别的一种常用方法是有 T. Schreiber 提出的基于错误函数的估计方法[83]。该方法是基于非线性动力学技术的一种噪声级别估计算法,算法利用非线性动力系统在相空间中的轨迹演化特征受到噪声干扰后产生的变化量(与预期轨道的偏差)来判别噪声的级别。该算法假设噪声为:

1) 随机且不相关的波动。

2) 均值为零且方差固定。

3) 噪声呈高斯分布。

具体算法如下[83,84]:

设观测时间序列为 $\{x_n\}_{n=1}^{N}$,作式(2.14)的相空间重构,按 2.3.1 节的方法计算关联维数 d,对于计算关联维数 d 的曲线中存在的一些波动,T. Schreiber 等认为是由于噪声引起的,因此可以认为关联维数和搜索半径参数 r 之间存在函数关系,即 $d=d(r)$。

假设 k 是实际的嵌入维数,且 $k>m$,则有如下估计噪声级的公式

$$\frac{d_k(r)-d_m(r)}{k-m}=g\left(\frac{r}{2\sigma}\right),\ g(z)=\frac{2}{\sqrt{\pi}}\frac{z\mathrm{e}^{-z^2}}{\mathrm{erf}(z)} \tag{2.101}$$

式中,erf(·)为错误函数;σ 为噪声级;$d_k(r)$和 $d_m(r)$分别为嵌入维数为 k 和 m 时计算的关联维数。通过 k 取不同值对式(2.101)进行拟合,可估计噪声级 σ。

该方法通过$\frac{d_k(r)-d_m(r)}{k-m}$揭示了噪声的分布,尽管这个方法不十分严格,但还是在实际应用中展示了它的有效性。由于该算法不能提供$\frac{d_k(r)-d_m(r)}{k-m}$和错误函数的精确拟合,大多数情况下通过下面的办法来计算噪声级 σ。

对一系列 $r_i(i=1,2,\cdots,I)$和 $k_j>m(j=1,2,\cdots,J)$,噪声级 σ 可以取使函数

$$\sum_{j=1}^{J}\sum_{i=1}^{I}\left[\frac{d_{k_j}(r_i)-d_m(r_i)}{k_i-m}-\frac{r_i\exp\left(-\frac{r_i^2}{4\sigma^2}\right)}{\sigma\sqrt{\pi}\,\mathrm{erf}\left(\frac{r_i}{2\sigma}\right)}\right] \tag{2.102}$$

达到最小时的 σ 值。

随着非线性科学技术的发展,特别是基于递归图的定量分析方法逐步成熟,K. Urbanowicz 和 J. A. Holyst 等在 2003 年提出了一种基于熵的噪声估计方法,

将 T. Schreiber 等的基于错误函数的噪声估计方法和递归图的定量分析技术利用熵作为桥梁关联在一起，称为粗糙纹理型熵估计方法，这种方法特别针对高噪声级别的估计而设计，具体算法可见文献[85]。另外，文献[86]还给出了一种基于高斯核权函数的噪声级别估计方法，该方法在估计噪声级别准确度上有一定程度的改进。

2.7.2 噪声的降低

降低噪声一直是信息处理领域的一个非常重要的课题，它本质上是通过滤波来实现的。经过多年的研究和应用实践，人们已经获得许多成果，文献[87～91]对非线性时间序列中的噪声处理方法进行了详细的说明。在金融领域，常常采用移动平均法来对时间序列进行平滑处理，它忽略了一些小的波动而反映序列的主要趋势，因而本质上也是一种滤波技术。但是，这种方法非常粗略，在降噪声的同时，把许多有用信息也一并去掉了。下面将简单说明传统滤波方法及其缺陷并介绍基于非线性动力学的非线性时间序列噪声降低的常用方法。

1. 传统滤波方法及其缺陷

传统的滤波理论是假定观测时间序列中的有用数据和噪声数据具有不同的频谱，即有用数据和噪声数据是频谱分离的，从而可以设计一种理想的滤波器 $H(\omega)$ 进行噪声滤波。显然这种滤波器没有考虑系统的其他特性，例如系统的运行（演化）规律、信号与噪声的统计特征，因而这种方法存在较多的缺陷。现代滤波理论中的一个典型代表——维纳滤波理论提出：寻找线性滤波器的最佳冲击响应或传递函数，使得滤波器的输出波形为输入波形的最佳估计。虽然维纳滤波思想比标准滤波器有了明显的改进，但这种方法对时间序列的平稳性要求使得它不适用于许多时间序列的分析和处理。卡尔曼滤波器改以最小方差为估计标准来寻找一套递推估计，这种方法适用于线性时变动力系统产生的非平稳时间序列，但这种方法要求知道动力系统的运动规律以建立准确的状态方程，同时要求噪声的统计特性作为一个关键输入参数。对于实际问题中复杂系统的观测时间序列而言这两个要求都是无法满足的。

另外一类常用的噪声处理技术就是小波方法，其中主要有两种方法：模极大值法和非线性阈值法。小波降噪是通过对信号进行多分辨率分解，可从多个层次上分析和考察信号的特征。由于非平稳信号有许多尖峰和突变，这些尖峰和突变往往具有较高的频率，采用传统的滤波方法不能将真实信号（有用信号）的高频干扰和噪声引起的高频干扰加以区分。若低通滤波带宽太窄，则将一部分有用信号去掉了，引起信号失真，不利于进一步的数据分析和预测，而带宽太大，又达不到降噪的目的。利用小波对信号进行多分辨率处理时，实际上相当于对信号同时进行低通和高通滤波，其低频系数主要反映真实信号，而高频系数主要反映噪声和真实信号的细节信息，但信号的细节信息与噪声在高频分解系数中会表现出一些不同的特性，这

正是利用小波进行降噪的切入点。事实上，对复杂系统的观测时间序列来说，很难给噪声一个准确的定义。通常那些随机性强、无法从中找出有价值的规律的信号都可视为噪声信号。按这样的观点，任意形式分布的信号都可能归为噪声。但问题是在对时间序列进行处理之前，我们难以辨别哪些信号会对系统的确定性表现有贡献，而哪些会起干扰。因此，按保守的观点，一般将噪声定义为高斯白噪声形式。这样一个假设对实际复杂系统的观测时间序列的分析和预测而言是有诸多问题的，其中不仅仅包含外部测量噪声，而且有很多噪声又作为输入反馈进入动力系统成为内在噪声。

通过以上对传统的滤波技术分析，可以明确这类方法对实际复杂系统的观测时间序列的降噪是不合适的，从而需要采用基于非线性动力学技术的噪声降低技术来处理实际复杂系统的观测时间序列中的噪声。这类方法主要是通过重构非线性动力系统的相空间，研究吸引子的演化轨迹特征，从而进行噪声降低。

2. 简单非线性降噪法

简单非线性降噪法是一种非常简单的非线性时间序列降噪方法，具有计算复杂程度低、计算速度快的优点[92]。其核心思想是：使用观测时间序列的局部平均值代替时间序列的中间值，这是基于该时间序列由一个连续动力系统产生的这一假设。

噪声降低的过程非常简单，首先将观测时间序列嵌入到合适的空间，嵌入空间的维数必须高于其理论上需求的嵌入维数。对于每个嵌入向量 x_n，在相空间中找出 x_n 的 ε 邻域 $u_\varepsilon(x_n)=\{x_j:\ \|x_j-x_n\|<\varepsilon\}$，邻域的搜索半径 ε 应大于噪声的级别但应小于吸引子的典型曲率半径。由于这两个条件很难同时满足，因此必须根据处理结果进行反复调整。如果噪声级别小于吸引子典型曲率半径，则可以使搜索半径为噪声级别的 2 到 3 倍。

对于嵌入向量 $x_n=(x_{n-m},x_{n-m+1},\cdots,x_{n-1})$（这里设 $\tau=1$），修正过的中间值 $\hat{x}_{n-m/2}$ 可以通过计算 $u_\varepsilon(x_n)$ 的平均值得到，即

$$\hat{x}_{n-m/2}=\frac{1}{|u_\varepsilon(x_n)|}\sum_{x_j\in u_\varepsilon(x_n)}x_{j-m/2} \tag{2.103}$$

在计算完成后，所有参与过计算的 x_n 都被替代为 $\hat{x}_n$。重新调整搜索半径，再进行下次迭代，这样就可得到降噪后的时间序列 $\{\hat{x}_n\}_{n=1}^{N}$。

为了检验算法的有效性，对例 2.1 的 Henon 映射进行仿真计算。图 2.15 是干净的 Henon 映射吸引子，图 2.16 是施加 40% 的 Gauss 白噪声的 Henon 映射吸引子，图 2.17 是通过上述方法降噪后的 Henon 映射吸引子。

可见，图 2.17 的吸引子与原动力系统产生的干净时间序列经重构所得的吸引子整体形状基本相同，部分细节特征仍可保留。很明显，对于受到高级别噪声污染的非线性时间序列，该算法可以有效恢复并保持原有动力系统的特性。

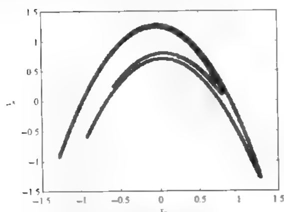


图 2.15　干净的 Henon 的吸引子

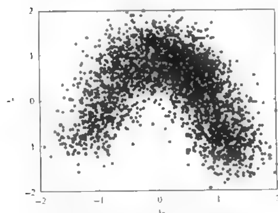


图 2.16　受到噪声干扰的 Henon 吸引子

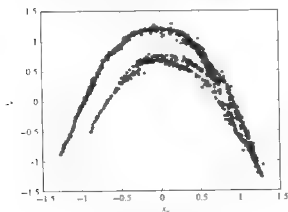


图 2.17　降噪后的 Henon 吸引子

第三章　基于多变量时间序列的系统非线性性检验

实际问题中的复杂系统往往是由多个变量描述的，通过观测或实验可获取多变量时间序列，例如从一个研究的系统中同时记录几个不同的变量，像气象学中的大气压、温度和湿度，生理学中的心率、血压、呼吸和血氧饱和度，经济学中一些货币相互之间的汇率、证券市场的分类指数时间序列等；或者在空间扩展系统中，从不同空间位置获得多道记录，像湍流、卫星数据、心电图和脑电图等。对这些多变量时间序列，同样需要研究以下问题：通过观测或实验获得的时间序列是来自于线性系统还是非线性系统？

对单变量时间序列，这一问题已在 2.5 节中进行过较详细的讨论。对多变量时间序列，文献[93～95]利用信息论中的熵定义线性冗余和冗余，线性冗余与冗余之间的差异可用于定性地检验实测时间序列是否来自于非线性系统；也可把冗余作为判别统计量，利用 2.5 节中的替代数据法来定量地检验实测时间序列是否来自于非线性系统。与 2.5 节中的相关方法相比，这种方法至少有三大优点：第一，为了检验变量之间各种特殊的依赖性，包括条件依赖性，可定义各种不同类型的冗余；第二，冗余可作为时间延迟间隔的函数，因此研究的依赖性不是基于统计的思想而是基于动力系统的思想；第三，对应于各种冗余，它们的线性冗余也存在，这些线性冗余可以用于检验生成的替代数据的质量，这样可避免由于生成的替代数据的不完美造成的非线性特性的虚假检验。

文献[93～95]中提出的方法在实际应用中的困难在于对实测的多变量时间序列，利用概率计算其冗余非常麻烦，我们把这一方法推广到广义冗余，建立了广义冗余与广义关联积分之间的关系，从而可用广义关联积分来计算广义冗余，计算比较简便，可操作性较强。在此基础上建立了定性和定量检验多变量时间序列是否来自于非线性系统的方法[96～99]。

3.1　随机变量的线性冗余和广义冗余

设随机变量 $X_1, X_2, \cdots, X_L$ 的概率密度分别为 $p_1(x_1), p_2(x_2), \cdots, p_L(x_L)$，随机向量$(X_1, X_2, \cdots, X_L)$的概率密度为 $p(x_1, x_2, \cdots, x_L)$。定义离散随机变量 X_l 的 q 阶 Renyi 熵为[100]

$$H_q(X_l) = \frac{1}{1-q}\log_2 \sum_{x_l} [p_l(x_l)]^q,\ l = 1, 2, \cdots, L;\quad q \neq 1 \qquad (3.1)$$

$$H_1(X_l) = -\sum_{x_l} p_l(x_l)\log_2[p_l(x_l)],\ l = 1,2,\cdots,L \tag{3.2}$$

定义随机变量 $X_1,X_2,\cdots,X_L$ 的联合 q 阶 Renyi 熵为[100]

$$H_q(X_1,X_2,\cdots,X_L) = \frac{1}{1-q}\log_2\sum_{x_1x_2\cdots x_L}[p(x_1,x_2,\cdots,x_L)]^q,\quad q\neq 1 \tag{3.3}$$

$$H_1(X_1,X_2,\cdots,X_L) = -\sum_{x_1x_2\cdots x_L} p(x_1,x_2,\cdots,x_L)\log_2 p(x_1,x_2,\cdots,x_L) \tag{3.4}$$

显然,式(3.2)和式(3.4)分别是式(3.1)和式(3.3))当 $q\to 1$ 时的极限,即 $q\to 1$ 时,Renyi 熵成为 Shannon 熵。

无论 q 取什么值,可定义定量刻画了包含在 $X_1,X_2,\cdots,X_L$ 中公共信息的平均数的 L 个变量 $X_1,X_2,\cdots,X_L$ 的广义冗余

$$R_q(X_1,X_2,\cdots,X_L) = H_q(X_1)+\cdots+H_q(X_L)-H_q(X_1,X_2,\cdots,X_L) \tag{3.5}$$

$q=1$ 时,即为文献[93~95]中的冗余。

定义定量刻画了变量 X_L 包含在变量 $X_1,X_2,\cdots,X_{L-1}$ 中信息的平均数的广义边缘冗余为

$$\begin{aligned}\rho_q(X_1,\cdots,X_{L-1},X_L) &= H_q(X_1,\cdots,X_{L-1}) + H_q(X_L) - H_q(X_1,X_2,\cdots,X_L)\\ &= R_q(X_1,\cdots,X_{L-1},X_L) - R_q(X_1,\cdots,X_{L-1})\end{aligned} \tag{3.6}$$

还可定义定量刻画变量间或变量群之间信息平均数的各种类型的广义冗余。例如,考虑变量 $X_1,\cdots,X_{L_1};Y_1,\cdots,Y_{L_2};Z_1,\cdots,Z_{L_3}$,则三个变量群之间的广义冗余为

$$\begin{aligned}&R_q(X_1,\cdots,X_{L_1};Y_1,\cdots,Y_{L_2};Z_1,\cdots,Z_{L_3})\\ &\quad = H_q(X_1,\cdots,X_{L_1}) + H_q(Y_1,\cdots,Y_{L_2}) + H_q(Z_1,\cdots,Z_{L_3})\\ &\qquad - H_q(X_1,\cdots,X_{L_1},Y_1,\cdots,Y_{L_2},Z_1,\cdots,Z_{L_3})\end{aligned} \tag{3.7}$$

或者定义变量群 X 和 Y 一起与 Z 之间的广义冗余(可看作是广义边缘冗余在群变量之间的推广)为

$$\begin{aligned}&R_q(X_1,\cdots,X_{L_1};Y_1,\cdots,Y_{L_2};Z_1,\cdots,Z_{L_3})\\ &\quad = H_q(X_1,\cdots,X_{L_1};Y_1,\cdots,Y_{L_2}) + H_q(Z_1,\cdots,Z_{L_3})\\ &\qquad - H_q(X_1,\cdots,X_{L_1};Y_1,\cdots,Y_{L_2};Z_1,\cdots,Z_{L_3})\end{aligned} \tag{3.8}$$

设 $X_1,X_2,\cdots,X_L$ 为 L 维随机变量,具有正态分布,均值为零,协方差矩阵为 $C=(c_{ij})_{L\times L}$,记 ξ 为 C 的逆矩阵,ξ_{ij} 是 ξ 的元素,$|\xi|$ 表示方阵 ξ 的行列式,则多变量 Gaussian 分布的密度分布函数为[100]

$$p(x_1,x_2,\cdots,x_L) = \frac{|\xi|^{1/2}}{(2\pi)^{1/2}}\exp\left(-\frac{1}{2}\sum_{i,j=1}^{L}\xi_{ij}x_ix_j\right)$$

按前面的定义计算其联合 q 阶 Renyi 熵及联合 Shannon 熵,得

$$H_q(X_1,X_2,\cdots,X_L)=\frac{1}{1-q}\log_2\left[\frac{|\xi|^{q/2}}{(2\pi)^{Lq/2}}\int\cdots\int\exp\left(-\frac{q}{2}\sum_{i,j=1}^{L}\xi_{ij}x_ix_j\right)\mathrm{d}x_1\cdots\mathrm{d}x_L\right]$$

$$=\frac{1}{1-q}\log_2\left[\frac{|\xi|^{(q-1)/2}}{q^{L/2}(2\pi)^{(q-1)L/2}}\right]$$

$$-\frac{1}{2}L\log_2(2\pi)+\frac{1}{2}\log_2|C|+\frac{L\log_2 q}{2(q-1)},\ q\neq 1 \qquad (3.9)$$

$$H_1(X_1,X_2,\cdots,X_L)=\lim_{q\to 1}H_q(X_1,X_2,\cdots,X_L)$$

$$=\frac{1}{2}L\log_2(2\pi e)+\frac{1}{2}\log_2|C| \qquad (3.10)$$

因此无论 q 取什么值，其广义冗余都为

$$R_q(X_1,X_2,\cdots,X_L)=\frac{1}{2}\sum_{i=1}^{L}\log_2(c_{ii})-\frac{1}{2}\log_2|C| \qquad (3.11)$$

其中 $|C|=\prod_{j=1}^{L}\lambda_j$，$\lambda_j$ 是 C 的特征值。可以发现，多变量 Gauss 分布的广义冗余与 q 无关，为此，我们定义任意随机变量 $X_1,X_2,\cdots,X_L$ 的线性冗余为

$$L(X_1,X_2,\cdots,X_L)=\frac{1}{2}\sum_{i=1}^{l}\log_2(c_{ii})-\frac{1}{2}\log_2|C| \qquad (3.12)$$

其中 $|C|=\prod_{j=1}^{L}\lambda_j$，$\lambda_j$ 是 C 的特征值，C 为协方差矩阵。

当随机变量 $X_1,X_2,\cdots,X_L$ 为 L 维 Gauss 分布时，广义冗余与线性冗余是一致的。特别当 $X_1,X_2,\cdots,X_L$ 具有均值零，方差 1，C 为其关联矩阵，λ_j，$j=1,2,\cdots,L$ 为 C 的特征值时，

$$L(X_1,X_2,\cdots,X_L)=\frac{1}{2}\sum_{j=1}^{L}\log_2\lambda_j \qquad (3.13)$$

类似地，定义定量刻画 X_L 在 $X_1,X_2,\cdots,X_{L-1}$ 中的线性依赖性的线性边缘冗余为

$$\lambda(X_1,\cdots,X_{L-1},X_L)=L(X_1,\cdots,X_{L-1},X_L)-L(X_1,\cdots,X_{L-1}) \qquad (3.14)$$

广义冗余检验了研究数据的所有依赖性，而线性冗余只考虑线性依赖性。

3.2　时间序列广义冗余的计算

实测多变量时间序列 $\{x_{l,n}\}_{n=1}^{N}(l=1,2,\cdots,L)$ 可以认为是多变量平稳遍历随机过程 $X_1,X_2,\cdots,X_L$ 的实现，则 $p_l(x_l)=\mathrm{prob}\{X_l=(x_{l,1},x_{l,2},\cdots,x_{l,N})\}(l=1,2,\cdots,L)$，$p(x_1,x_2,\cdots,x_L)=\mathrm{prob}\{X_1=(x_{1,1},x_{1,2},\cdots,x_{1,N}),X_2=(x_{2,1},x_{2,2},\cdots,x_{2,N}),\cdots,X_L=(x_{L,1},x_{L,2},\cdots,x_{L,N})\}$。由于遍历性，所有的信息论函数可以由时间平均代替总体平均进行估计。为了便于计算，把多变量时间序列 $\{x_{l,n}\}_{n=1}^{N}(l=1,2,\cdots,L)$ 每一变量分别标准化，即令

$$x_{l,n}^{*}=\frac{x_{l,n}-\frac{1}{N}\sum_{n=1}^{N}x_{l,n}}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(x_{l,n}-\frac{1}{N}\sum_{n=1}^{N}x_{l,n}\right)^{2}}},\ l=1,2,\cdots,L;\ n=1,2,\cdots,N \tag{3.15}$$

则$\{x_{l,n}^{*}\}_{n=1}^{N}(l=1,2,\cdots,L)$的均值为零，方差为1。这样线性冗余可由式(3.13)直接计算，其中关联矩阵的特征值可由文献[101]中的SVDCMP程序计算。

实测多变量时间序列的广义冗余可由以下两种方法计算。

3.2.1　直方图及盒计数法

为了计算$H_q(X_l)$，对时间序列$\{x_{l,n}\}_{n=1}^{N}(l=1,2,\cdots,L)$从$\min\{x_{l,n}:n=1,2,\cdots,N\}$到$\max\{x_{l,n}:n=1,2,\cdots,N\}$作$K$等分，此时计算精度为

$$\delta_l=\frac{\max\{x_{l,n}:n=1,2,\cdots,N\}-\min\{x_{l,n}:n=1,2,\cdots,N\}}{K} \tag{3.16}$$

利用直方图可得每个小区间内的概率$p_l(k)(k=1,2,\cdots,K)$，则

$$\begin{aligned}H_q(X_l)&=\frac{1}{1-q}\log_2\sum_{x_l}[p_l(x_l)]^q\\&=\frac{1}{1-q}\log_2\sum_{k=1}^{K}[p_l(k)]^q,\ l=1,2,\cdots,L\end{aligned} \tag{3.17}$$

为了计算$H_q(X_1,X_2,\cdots,X_L)$，最直接的方法是利用盒计数法，即L维空间被分成一系列大小为δ的超立方体(δ为直径)，设为K个，计算第k个小立方体内的点数n_k，则概率$p_k=\frac{n_k}{N}(k=2,\cdots,K)$，因此

$$\begin{aligned}H_q(X_1,X_2,\cdots,X_L)&=\frac{1}{1-q}\log_2\sum_{x_1x_2\cdots x_L}[p(x_1,x_2,\cdots,x_L)]^q\\&=\frac{1}{1-q}\log_2\sum_{k=1}^{K}(p_k)^q\\&=\frac{1}{1-q}\log_2\sum_{k=1}^{K}\left(\frac{n_k}{N}\right)^q\end{aligned} \tag{3.18}$$

文献[95]在$q\to1$的情况下提出了依赖于局部密度的改进算法，但这些方法在实现时都比较复杂，且费时。

3.2.2　关联积分计算法

记L维向量$\boldsymbol{x}_i=(x_{1,i},x_{2,i},\cdots,x_{L,i})$与$\boldsymbol{x}_j=(x_{1,j},x_{2,j},\cdots,x_{L,j})$之间的距离为$\|\boldsymbol{x}_i-\boldsymbol{x}_j\|=\max\{|x_{l,i}-x_{l,j}|:1\leqslant l\leqslant L\}$（也可以取欧氏范数），则$\sum_{j=1,j\neq i}^{N}H(r-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|)$表示除$\boldsymbol{x}_i$本身外到$\boldsymbol{x}_i$的距离小于$r$的$\boldsymbol{x}_j$的点数，其中

$H(\cdot)$是 Heavside 函数，定义 L 变量时间序列$\{x_n\}_{n=1}^N$的 q 阶广义关联积分为

$$C_q^L(r) = \left\{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N-1}\sum_{j=1,j\neq i}^{N}H(r - \|x_i - x_j\|)\right]^{q-1}\right\}^{\frac{1}{q-1}} \tag{3.19}$$

当 $q=2$ 时就是通常的关联积分，$q\to1$ 时，即为一阶广义关联积分，即

$$\begin{aligned}\log_2 C_1^L(r) &= \log_2 \lim_{q\to1} C_q^L(r)\\ &= \lim_{q\to1}\frac{1}{q-1}\log_2\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N-1}\sum_{j=1,j\neq i}^{N}H(r - \|x_i - x_j\|)\right]^{q-1}\\ &= \frac{1}{N}\sum_{i=1}^{N}\log_2\frac{1}{N-1}\sum_{j=1,j\neq i}^{N}H(r - \|x_i - x_j\|)\end{aligned} \tag{3.20}$$

记 $p_i(x_{1,i},x_{2,i},\cdots,x_{L,i};r)(i=1,2,\cdots,N)$为中心在$(x_{1,i},x_{2,i},\cdots,x_{L,i})$点处，半径为 $r-\dfrac{\delta}{2}$的球内的概率，利用最大范数可得

$$\begin{aligned}\sum_{x_1x_2\cdots x_L}[p(x_1,x_2,\cdots,x_L)]^q &\approx \frac{1}{N}\sum_{i=1}^{N}[p_i(x_{1,i},x_{2,i},\cdots,x_{L,i};r)]^{q-1}\\ &= \frac{1}{N}\sum_{i=1}^{N}\left[\frac{n_i}{N}\right]^{q-1}\\ &\approx \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N-1}\sum_{j=1,j\neq i}^{N}H(r-\|x_i-x_j\|)\right]^{q-1}\end{aligned} \tag{3.21}$$

因此，当 $q\neq1$ 时

$$\begin{aligned}H_q(X_1,X_2,\cdots,X_L) &= \frac{1}{1-q}\log_2\sum_{x_1x_2\cdots x_L}[p(x_1,x_2,\cdots,x_L)]^q\\ &\approx \frac{1}{1-q}\log_2\left\{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N-1}\sum_{j=1,j\neq i}^{N}H(r - \|x_i - x_j\|)\right]^{q-1}\right\}\\ &= -\log_2 C_q^L(r)\end{aligned} \tag{3.22}$$

当 $q=1$ 时，在式(3.21)两边取对数，然后同除 $q-1$，令 $q\to1$，得

$$\begin{aligned}H_1(X_1,X_2,\cdots,X_L) &= -\sum_{x_1x_2\cdots x_L}p(x_1,x_2,\cdots,x_L)\log_2 p(x_1,x_2,\cdots,x_L)\\ &\approx -\frac{1}{N}\sum_{i=1}^{N}\log_2\left[\frac{1}{N-1}\sum_{j=1,j\neq i}^{N}H(r - \|x_i - x_j\|)\right]\\ &= -\log_2 C_1^L(r)\end{aligned} \tag{3.23}$$

同理，无论 q 取什么值，都有

$$H_q(X_l) = -\log_2 C_{q,l}^1(r) \tag{3.24}$$

其中 $C_{q,l}^1(r)$是单变量时间序列$\{x_{l,n}\}_{n=1}^N$的 q 阶广义积分，$l=1,2,\cdots,L$。

要注意的是，这里单变量时间序列没有作延迟重构，即当 $l=1,2,\cdots,L$ 时

$$C_{q,l}^{1}(r)=\left\{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N-1}\sum_{j=1,j\neq i}^{N}H(r-|x_{l,i}-x_{l,j}|)\right]^{q-1}\right\}^{\frac{1}{q-1}},\quad q\neq 1 \tag{3.25}$$

$$\log_2 C_{1,l}^{1}(r)=\frac{1}{N}\sum_{i=1}^{N}\log_2\frac{1}{N-1}\sum_{j=1,j\neq i}^{N}H(r-|x_{l,i}-x_{l,j}|) \tag{3.26}$$

这样,可得用广义关联积分表示的广义冗余和广义边缘冗余,即

$$\begin{aligned}R_q(X_1,X_2,\cdots,X_L)&=H_q(X_1)+\cdots+H_q(X_L)-H_q(X_1,X_2,\cdots,X_L)\\&=\log_2\frac{C_q^L(r)}{C_{q,1}^1(r)\cdots C_{q,L}^1(r)}\end{aligned} \tag{3.27}$$

$$\begin{aligned}\rho_q(X_1,\cdots,X_{L-1},X_L)&=R_q(X_1,\cdots,X_{L-1},X_L)-R_q(X_1,\cdots,X_{L-1})\\&=\log_2\frac{C_q^L(r)}{C_q^{L-1}(r)C_{q,L}^1(r)}\end{aligned} \tag{3.28}$$

3.3 系统非线性性的定性和定量检验

对实测多变量时间序列$\{x_{l,n}\}_{n=1}^{N}$,$l=1,2,\cdots,L$,研究延迟时间序列之间的依赖性是可能的和有用的。假设时间序列是平稳的,则广义冗余$R_q(x_{1,n};x_{2,n+\tau_1};\cdots;x_{L,n+\tau_{L-1}})$不依赖于$n$,只是延迟时间$\tau_1,\tau_2,\cdots,\tau_{L-1}$的函数,因此可得广义冗余和广义边缘冗余

$$R_q(\tau_1,\tau_2,\cdots,\tau_{L-1})=R_q(x_{1,n};x_{2,n+\tau_1};\cdots;x_{L,n+\tau_{L-1}}) \tag{3.29}$$

$$\rho_q(\tau_1,\tau_2,\cdots,\tau_{L-1})=R_q(\tau_1,\tau_2,\cdots,\tau_{L-1})-R_q(\tau_1,\tau_2,\cdots,\tau_{L-2}) \tag{3.30}$$

同理,也可得多变量时间序列$\{x_{l,n}\}_{n=1}^{N}$,$l=1,2,\cdots,L$的线性冗余和线性边缘冗余

$$L(\tau_1,\tau_2,\cdots,\tau_{L-1})=L(x_{1,n};x_{2,n+\tau_1};\cdots;x_{L,n+\tau_{L-1}}) \tag{3.31}$$

$$\lambda(\tau_1,\tau_2,\cdots,\tau_{L-1})=L(\tau_1,\tau_2,\cdots,\tau_{L-1})-L(\tau_1,\tau_2,\cdots,\tau_{L-2}) \tag{3.32}$$

这一方法同样可用于单变量时间序列,对实测单变量时间序列$\{x_n\}_{n=1}^{N}$,通过相空间重构为m维向量$\boldsymbol{x}_n=(x_n,x_{n-\tau},\cdots,x_{n-(m-1)\tau})\in R^m(n=N_0,N_0+1,\cdots,N)$,其中$m,\tau$为正整数,若把$\{x_n\},\{x_{n-\tau}\},\cdots,\{x_{n-(m-1)\tau}\}$分别看作是随机变量$X_1,X_2,\cdots,X_m$的实现,则在假设时间序列是平稳的条件下,$R_q(x_n,x_{n-\tau},\cdots,x_{n-(m-1)\tau})=R_q(X_1,X_2,\cdots,X_m)$与$n$无关,只与$m$和$\tau$有关,因此可得单变量时间序列$\{x_n\}_{n=1}^{N}$的广义冗余和广义边缘冗余

$$R_q^m(\tau)=R_q(x_n,x_{n-\tau},\cdots,x_{n-(m-1)\tau}) \tag{3.33}$$

$$\rho_q^m(\tau)=R_q^m(\tau)-R_q^{m-1}(\tau) \tag{3.34}$$

同理,可得单变量时间序列$\{x_n\}_{n=1}^{N}$的线性冗余和线性边缘冗余

$$L^m(\tau)=L(x_n,x_{n-\tau},\cdots,x_{n-(m-1)\tau}) \tag{3.35}$$

$$\lambda^m(\tau) = L^m(\tau) - L^{m-1}(\tau) \tag{3.36}$$

3.3.1 系统非线性性的定性检验

根据定义，线性冗余只刻画了研究的时间序列的线性结构，而广义冗余刻画了研究的时间序列的非线性结构，因此可比较广义冗余 $R_q(\tau_1,\tau_2,\cdots,\tau_{L-1})$[或$R_q^m(\tau)$]和线性冗余 $L(\tau_1,\tau_2,\cdots,\tau_{L-1})$[或 $L^m(\tau)$]关于 $\tau_1,\tau_2,\cdots,\tau_{L-1}$(或 τ)的图形。如果它们的形状相同或非常类似，则足以认为研究的时间序列可用线性过程描述，而较大的差异表明不同变量的时间序列之间(或时间序列与它的延迟时间序列之间)具有非线性关系。这一过程可以用图 3.1 表示。

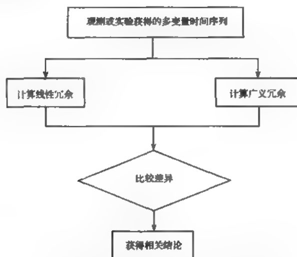


图 3.1 利用冗余定性检验系统非线性性的流程图

值得注意的是，要比较的是 $R_q(\tau_1,\tau_2,\cdots,\tau_{L-1})$[或 $R_q^m(\tau)$]与 $L(\tau_1,\tau_2,\cdots,\tau_{L-1})$[或 $L^m(\tau)$]关于 $\tau_1,\tau_2,\cdots,\tau_{L-1}$(或 τ)的函数的形状，而不是它们特殊的数值，这是由于它们的不同数值性质所致。对固定的 m 和 $\tau_1,\tau_2,\cdots,\tau_{L-1}$(或 τ)，$L(\tau_1,\tau_2,\cdots,\tau_{L-1})$[或 $L^m(\tau)$]只与时间序列的长度有关，而 $R_q(\tau_1,\tau_2,\cdots,\tau_{L-1})$[或 $R_q^m(\tau)$]依赖于使用的数值程序(状态空间剖分越精细，估计的 $R_q(\tau_1,\tau_2,\cdots,\tau_{L-1})$[或 $R_q^m(\tau)$]值越大)，因此不能用任何正式的统计程序定量地确定广义冗余之间是否有显著差异，但 $R_q(\tau_1,\tau_2,\cdots,\tau_{L-1})$[或 $R_q^m(\tau)$]的形状对不同的剖分是一样的，因此定性的比较是可能的，由此可检验观测或实验获得的时间序列是来自于线性系统还是非线性系统。实际应用中，对多变量实测时间序列，可计算 $R_q(\tau_1,\tau_2,\cdots,\tau_{L-1})$及 $L(\tau_1,\tau_2,\cdots,\tau_{L-1})$的一些一维截面，即取其中一个为 τ，其余为零。

3.3.2　系统非线性性的定量检验

为了定量地研究观测时间序列是否来自非线性系统，可以把线性冗余和广义冗余作为检验统计量，利用 2.5 节的替代数据法进行检验。

由于在实际应用中利用 Fourier 变换生成的替代时间序列不总是保持观测时间序列的线性性质，因此，只有观测时间序列和替代时间序列的广义冗余有明显差异，而它们的线性冗余没有明显差异时，认为观测时间序列来自于非线性系统才比较合理。因此，同时用线性冗余和广义冗余作为检验统计量比其他检验统计量具有更好的效果。

利用线性冗余和广义冗余检验系统非线性性的过程如图 3.2 所示。

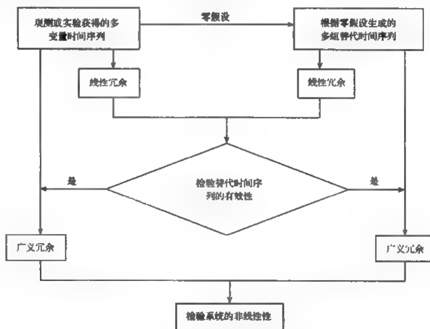


图 3.2　利用冗余定量检验系统非线性性的流程图

设 $\{x_{l,n}\}_{n=1}^{N}(l=1,2,\cdots,L)$ 为标准化后的观测时间序列，对每个变量分别做以下离散 Fourier 变换

$$y_{l,k}=\sum_{n=1}^{N}x_{l,n}\exp\left(-\mathrm{i}\,\frac{2\pi(k-1)(n-1)}{N}\right),\ k=1,2,\cdots,N;\ l=1,2,\cdots,L \tag{3.37}$$

其中 $i=\sqrt{-1}$。在区间 $[-\pi,\pi]$ 内取 N 个随机数 $\alpha_k(k=1,2,\cdots,N)$，并使其满足

$$\alpha_1=\alpha_{\frac{N}{2}+1}=0 \tag{3.38}$$

$$\alpha_k = -\alpha_{N+2-k},\ k = 2,3,\cdots,\frac{N}{2} \tag{3.39}$$

然后做以下离散 Fourier 逆变换

$$\tilde{x}_{l,n} = \frac{1}{N}\sum_{k=1}^{N} y_{l,k}\exp(\mathrm{i}\alpha_k)\exp\left(\mathrm{i}\,\frac{2\pi(k-1)(n-1)}{N}\right),\ n=1,2,\cdots,N;\ l=1,2,\cdots,L \tag{3.40}$$

则得 $\{x_{l,n}\}_{n=1}^{N}(l=1,2,\cdots,L)$ 的替代时间序列 $\{\tilde{x}_{l,n}\}_{n=1}^{N}(l=1,2,\cdots,L)$。

由文献[102]可知，这样生成的替代时间序列是线性 Gaussian 多维随机时间序列，但对每个 $l=1,2,\cdots,L$，$\{x_{l,n}\}_{n=1}^{N}$ 与 $\{\tilde{x}_{l,n}\}_{n=1}^{N}$ 有相同的幅度分布，从而有相同的自相关函数，因此替代时间序列保持了原时间序列的所有线性性质，但改变了原时间序列的非线性性。

根据需要，多变量时间序列的替代时间序列也可由 2.5.1 节中的其他方法(如 AAFT、IAAFT 等)生成。

记 $R_q(\tau_1,\tau_2,\cdots,\tau_{L-1})$ 和 $L(\tau_1,\tau_2,\cdots,\tau_{L-1})$ 分别为观测多变量时间序列 $\{x_{l,n}\}_{n=1}^{N}$ $(l=1,2,\cdots,L)$ 的广义冗余和线性冗余，$R_q^s(\tau_1,\tau_2,\cdots,\tau_{L-1})$ 和 $L^s(\tau_1,\tau_2,\cdots,\tau_{L-1})$ 为某一组替代时间序列 $\{\tilde{x}_{l,n}\}_{n=1}^{N}(l=1,2,\cdots,L)$ 的广义冗余和线性冗余，由替代时间序列的生成原理知它们是线性的，因此

$$R_q^s(\tau_1,\tau_2,\cdots,\tau_{L-1}) = L^s(\tau_1,\tau_2,\cdots,\tau_{L-1}) \tag{3.41}$$

由于替代时间序列和原时间序列有相同的自关联函数，则

$$L^s(\tau_1,\tau_2,\cdots,\tau_{L-1}) = L(\tau_1,\tau_2,\cdots,\tau_{L-1}) \tag{3.42}$$

这样

$$R_q(\tau_1,\tau_2,\cdots,\tau_{L-1}) = L(\tau_1,\tau_2,\cdots,\tau_{L-1}) \tag{3.43}$$

等价于

$$R_q(\tau_1,\tau_2,\cdots,\tau_{L-1}) = R_q^s(\tau_1,\tau_2,\cdots,\tau_{L-1}) \tag{3.44}$$

由于比较的量具有相同的数值性质，只要在估计 $R_q(\tau_1,\tau_2,\cdots,\tau_{L-1})$ 和 $R_q^s(\tau_1,\tau_2,\cdots,\tau_{L-1})$ 时用相同的数值参数，它们的值就可以定量地比较。具体应用中可定义如下广义冗余和线性冗余的显著性检验统计量：

$$\mathrm{DIFF}R_q = \frac{|R_q(\tau_1,\tau_2,\cdots,\tau_{L-1}) - \overline{R}_q^s(\tau_1,\tau_2,\cdots,\tau_{L-1})|}{\sigma_R(\tau_1,\tau_2,\cdots,\tau_{L-1})} \tag{3.45}$$

$$\mathrm{DIFF}L = \frac{|L(\tau_1,\tau_2,\cdots,\tau_{L-1}) - \overline{L}^s(\tau_1,\tau_2,\cdots,\tau_{L-1})|}{\sigma_L(\tau_1,\tau_2,\cdots,\tau_{L-1})} \tag{3.46}$$

式中，$\overline{R}_q^s(\tau_1,\tau_2,\cdots,\tau_{L-1})$ 和 $\overline{L}^s(\tau_1,\tau_2,\cdots,\tau_{L-1})$ 分别是多组替代时间序列广义冗余和线性冗余的均值；$\sigma_R(\tau_1,\tau_2,\cdots,\tau_{L-1})$ 和 $\sigma_L(\tau_1,\tau_2,\cdots,\tau_{L-1})$ 分别是它们的标准差。

如果广义冗余和线性冗余服从正态分布(数值实验表明这样的假设是合理的)，则对观测时间序列所作的零假设当显著性检验统计量的值大于 1.96 时可以以小于 5% 的显著水平被拒绝。

同理可定义线性边缘冗余和广义边缘冗余的显著性检验统计量。实际应用中，同定性检验类似，对多变量实测时间序列，可计算 $R_q(\tau_1,\tau_2,\cdots,\tau_{L-1})$ 及 $L(\tau_1,\tau_2,\cdots,\tau_{L-1})$ 的一些一维截面，即取其中一个为 τ，其余为零。

对单变量观测时间序列，根据零假设生成替代时间序列后，同样可以由广义冗余和线性冗余的显著性检验统计量：

$$\mathrm{DIFF}R_q^m(\tau)=\frac{|R_q^m(\tau)-\overline{R}_q^m(\tau)|}{\sigma_R(\tau)} \tag{3.47}$$

$$\mathrm{DIFF}L^m(\tau)=\frac{|L^m(\tau)-\overline{L}^m(\tau)|}{\sigma_L(\tau)} \tag{3.48}$$

式中，$R_q^m(\tau)$ 和 $L^m(\tau)$ 分别是观测时间序列的广义冗余和线性冗余，分别由式(3.33)和式(3.35)计算；$\overline{R}_q^m(\tau)$ 和 $\overline{L}^m(\tau)$ 分别是根据零假设生成的多组替代时间序列的广义冗余和线性冗余的均值；$\sigma_R(\tau)$ 和 $\sigma_L(\tau)$ 分别是它们的标准差。

由于在实际应用中利用 Fourier 变换生成的替代时间序列不总是保持原始数据的线性性质，因此，只有原时间序列和替代时间序列的线性冗余没有明显差异时，而它们的广义冗余产生明显的差异时拒绝零假设才不会产生由于替代时间序列的线性性质变化而引起的错误结果，但当观测时间序列和替代时间序列的线性冗余也产生显著差异时，不能理所当然地认为研究的系统具有非线性特征，因为这样的替代时间序列未能保持原时间序列的自关联函数。

3.4　仿真模拟

例 3.1　两变量线性自回归模型

$$\begin{cases} x_{1,n}=0.9x_{1,n-1}+\sigma_{1,n} \\ x_{2,n}=0.9x_{1,n-1}+x_{2,n-1}+\sigma_{2,n} \end{cases} \tag{3.49}$$

其中 $\sigma_{1,n}$，$\sigma_{2,n}$ 是均值为零，方差为 1 的随机时间序列。

取初始点 $x_{1,0}=0$，$x_{2,0}=0$，迭代生成长度 20 000 的时间序列。去掉前 3000 个点，从第 3001 个点开始取长度 $N=1024$ 的两变量时间序列 $\{x_{l,n}^*\}_{n=1}^N$，$l=1,2$ 作为研究对象，其时间历程图如图 3.3所示。

为了方便计算，先将两变量时间序列中每个变量时间序列分别标准化，即得到各变量均值为零，方差为 1 的两变量时间序列 $\{x_{l,n}\}_{n=1}^N$，$l=1,2$。

零假设 1：时间序列来自具常系数的高斯线性随机过程。

利用 2.5.1 节中随机相位化 Fourier 变换(FT)方法生产 19 组替代两变量时间序列，经计算，线性冗余的显著性检验统计量 $\mathrm{DIFF}L(x_{1,n};x_{2,n+\tau})$ 关于时间延迟间隔 τ 的函数如图 3.4(a)所示，广义冗余的显著性检验统计量 $\mathrm{DIFF}R_q(x_{1,n};x_{2,n+\tau})$ 关于时间延迟间隔 τ 的函数如图 3.4(b)($q=1$)、图 3.4(c)($q=2$)和图 3.4(d)($q=3$)所示。可见，线性冗余的显著性检验统计量的值小于 1.96，这表明

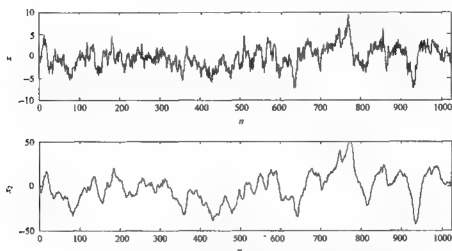


图 3.3 两变量线性自回归模型的时间历程图

原时间序列与替代时间序列的线性依赖结构没明显差异，即生成的替代时间序列是好的。同时广义冗余的显著性检验统计量的值几乎都小于 1.96，即不能有效地拒绝时间序列来自具常系数的高斯线性随机过程这一零假设，因此可以认为时间序列来自的系统不具有非线性特征，与实际情况一致。

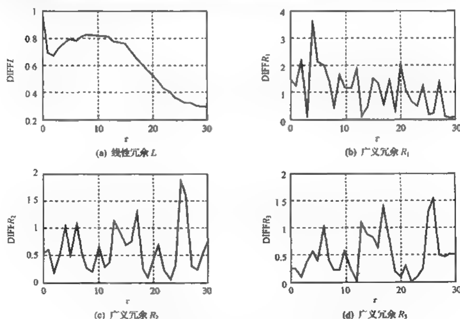


图 3.4 零假设 1 下线性冗余和广义冗余的显著性检验统计量

零假设 2：时间序列来自具常系数的高斯线性随机过程经过单调静态非线性变换。

利用 2.5.1 节中迭代的幅度调节 Fourier 变换(IAAFT)方法生产 19 组替代两变量时间序列，经计算，线性冗余的显著性检验统计量 DIFF$L(x_{1,n};x_{2,n+\tau})$关于时间延迟间隔 τ 的函数如图 3.5(a)所示，广义冗余的显著性检验统计量 DIFF$R_q(x_{1,n};x_{2,n+\tau})$关于时间延迟间隔 τ 的函数如图 3.5(b)($q=1$)、图 3.5(c)($q=2$)和图 3.5(d)($q=3$)所示。可见线性冗余的显著性检验统计量的值小于 1.96，这表明原时间序列与替代时间序列的线性依赖结构没明显差异，即生成的替代时间序列是好的。同时广义冗余的显著性检验统计量的值几乎都小于 1.96，即不能有效地拒绝时间序列来自具常系数的高斯线性随机过程经过静态非线性变换这一零假设，因此可以认为时间序列来自的系统不具有非线性特征，与实际情况一致。

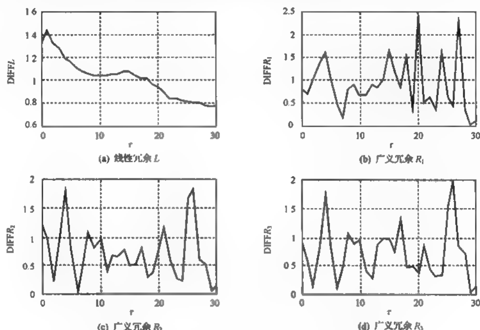


图 3.5　零假设 2 下线性冗余和广义冗余的显著性检验统计量

例 3.2　Lorenz 系统

$$\begin{cases} \dfrac{dx_1}{dt} = \sigma(x_2 - x_1) \\ \dfrac{dx_2}{dt} = x_1(r - x_3) - x_2 \\ \dfrac{dx_3}{dt} = x_1x_2 - bx_3 \end{cases} \tag{3.50}$$

其中 $\sigma=10$，$r=25$，$b=\dfrac{8}{3}$，取初值 $x_{1,0}=15.34$，$x_{2,0}=13.68$，$x_{3,0}=37.91$，积分步长

$h=0.04$，精度为 0.0001，利用四阶 Runge-Kute 法积分，迭代生成长度 20 000 的时间序列。去掉前 3000 个点，从第 3001 个点开始取长度 $N=1024$ 的三变量时间序列 $\{x_{l,n}^{*}\}_{n=1}^{N}(l=1,2,3)$ 作为研究对象，其时间历程图如图 3.6 所示。

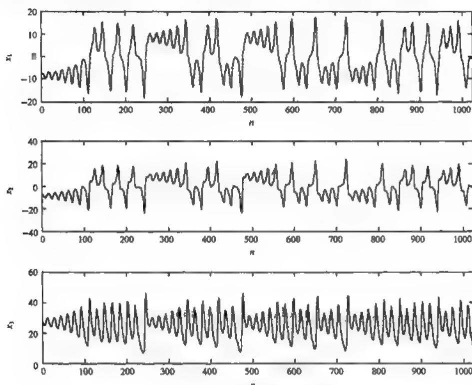


图 3.6 Lorenz 系统的时间历程图

为了方便计算，先将三变量原时间序列的每变量时间序列标准化，即得到各变量均值为零，方差为 1 的三变量时间序列 $\{x_{l,n}\}_{n=1}^{N}(l=1,2,3)$。

零假设 1：时间序列来自具常系数的高斯线性随机过程。

利用 2.5.1 节中随机相位化 Fourier 变换(FT)方法生产 19 组替代三变量时间序列，经计算，线性冗余的显著性检验统计量 DIFF$L(x_{1,n};x_{2,n+\tau};x_{3,n})$关于时间延迟间隔 τ 的函数如图 3.7(a)所示，广义冗余显著性检验统计量 DIFF$R_q(x_{1,n};x_{2,n+\tau};x_{3,n})$关于时间延迟间隔 τ 的函数如图 3.7(b)($q=1$)、图 3.7(c)($q=2$)和图 3.7(d)($q=3$)所示。

线性边缘冗余 $\lambda\{x_{1,n};x_{2,n+\tau};x_{3,n}\}$与广义边缘冗余 $\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n})$的显著性检验统计量 DIFF$\lambda(x_{1,n};x_{2,n+\tau};x_{3,n})$和 DIFF$\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n})$的计算结果如图 3.8所示。

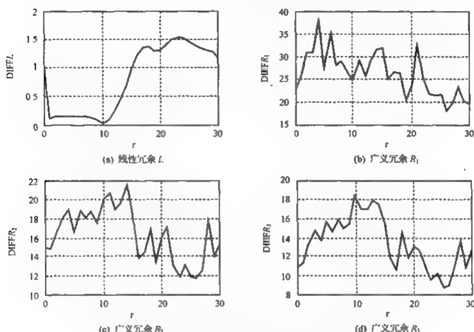


图 3.7 零假设 1 下线性冗余和广义冗余的显著性检验统计量

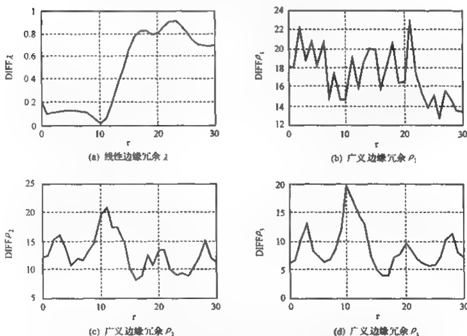


图 3.8 零假设 1 下线性边缘冗余和广义边缘冗余的显著性检验统计量

可见，线性冗余的显著性检验统计量的数值都小于1.96，这样原时间序列与替代数据的线性依赖结构没有明显差异，即替代时间序列在技术上是好的，不应该产生虚假的检验结果。但广义冗余的显著性检验统计量的数值较大，可以以95%的概率拒绝时间序列来自具常系数的高斯线性随机过程这一零假设，即可认为时间序列来自的系统具有非线性性，与实际的情况相吻合。

零假设2：时间序列来自具常系数的高斯线性随机过程经过单调静态非线性变换。

利用2.5.1节中迭代的幅度调节Fourier变换（IAAFT）方法生产19组替代三变量时间序列，经计算，线性冗余的显著性检验统计量 $\mathrm{DIFF}L(x_{1,n};x_{2,n+\tau};x_{3,n})$ 关于时间延迟间隔 τ 的函数如图3.9(a)所示，广义冗余显著性检验统计量 $\mathrm{DIFF}R_q(x_{1,n};x_{2,n+\tau};x_{3,n})$ 关于时间延迟间隔 τ 的函数如图3.9(b)($q=1$)、图3.9(c)($q=2$)和图3.9(d)($q=3$)所示。

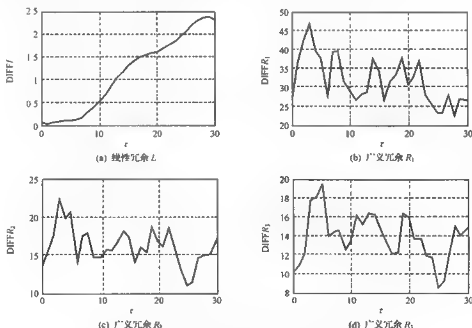


图3.9　零假设2下线性冗余和广义冗余的显著性检验统计量

线性边缘冗余 $\lambda(x_{1,n};x_{2,n+\tau};x_{3,n})$ 与广义边缘冗余 $\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n})$ 的显著性检验统计量 $\mathrm{DIFF}\lambda(x_{1,n};x_{2,n+\tau};x_{3,n})$ 和 $\mathrm{DIFF}\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n})$ 如图3.10所示。

可见线性冗余的显著性检验统计量的数值很小，而在 $24\leqslant\tau\leqslant30$ 线性冗余的显著性检验统计量出现大于2的现象，这可能与数据较少有关系。但广义冗余的显著性检验统计量的数值较大，可以以95%的概率拒绝时间序列来自具常系数的高

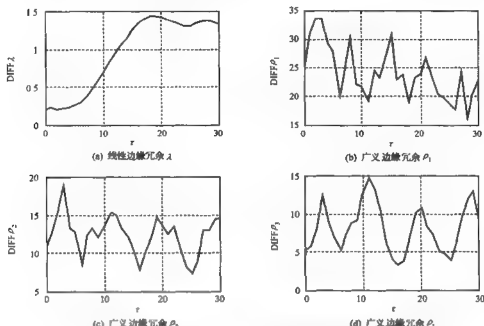


图 3.10　零假设 2 下线性边缘冗余和广义边缘冗余的显著性检验统计量

斯线性随机过程经过静态非线性变换这一零假设，即可认为时间序列来自的系统具有非线性性，与实际的情况相吻合。

例 3.3　白噪声时间序列。

人工生成长度为 20 000 的三个变量的白噪声时间序列，去掉每一变量的前面 3000 个点，从第 3001 个点选取长度 $N=1024$ 的三变量时间序列$\{x_{l,n}^{*}\}_{n=1}^{N}(l=1,2,3)$作为研究对象，其时间历程图如图 3.11 所示。

为了方便计算，先将三变量原时间序列的每变量时间序列标准化，即得到各变量均值为零，方差为 1 的多变量时间序列$\{x_{l,n}\}_{n=1}^{N}(l=1,2,3)$。

零假设 1：时间序列来自具常系数的高斯线性随机过程。

利用 2.5.1 节中随机相位化 Fourier 变换(FT)方法生产 19 组替代三变量时间序列，经计算，线性冗余的显著性检验统计量 $\mathrm{DIFF}L(x_{1,n};x_{2,n+\tau};x_{3,n})$关于时间延迟间隔 τ 的函数如图 3.12(a)所示，广义冗余显著性检验统计量 $\mathrm{DIFF}R_q(x_{1,n};x_{2,n+\tau};x_{3,n})$关于时间延迟间隔 τ 的函数见图 3.12(b)($q=1$)、图 3.12(c)($q=2$)和图 3.12(d)($q=3$)所示。

线性边缘冗余 $\lambda(x_{1,n};x_{2,n+\tau};x_{3,n})$与广义边缘冗余 $\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n})$的显著性检验统计量 $\mathrm{DIFF}\lambda(x_{1,n};x_{2,n+\tau};x_{3,n})$和 $\mathrm{DIFF}\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n})$的计算结果如图 3.13所示。

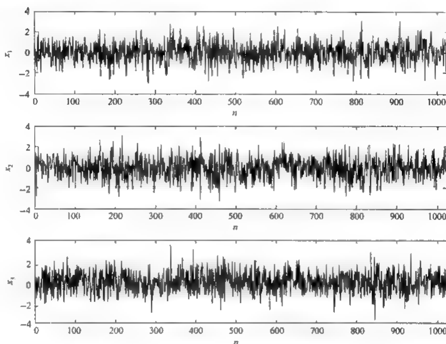


图 3.11　三变量的白噪声时间序列的时间历程图

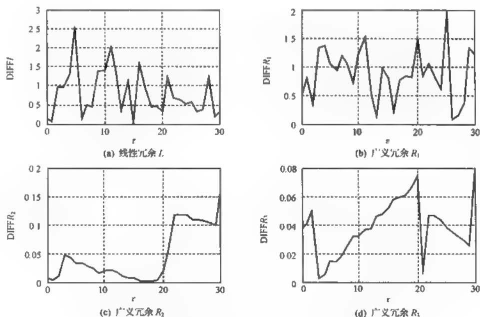


图 3.12　零假设 1 下线性冗余和广义冗余的显著性检验统计量

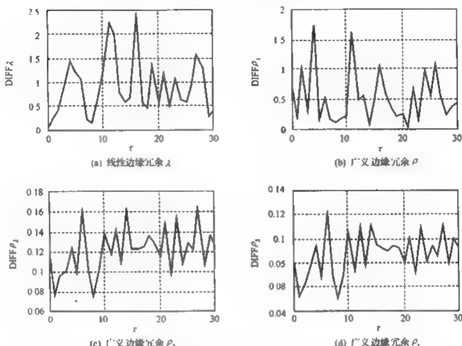


图 3.13　零假设 1 下线性边缘冗余和广义边缘冗余的显著性检验统计量

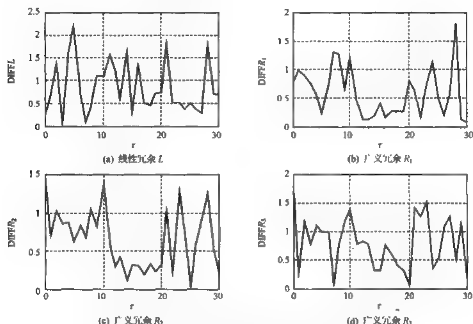
可见线性冗余的显著性检验统计量都很小，这表明原时间序列与替代时间序列的线性依赖结构没明显差异，即生成的替代时间序列是好的。同时广义冗余的显著性检验统计量的值几乎都小于 1.96，即不能有效地拒绝时间序列来自具常系数的高斯线性随机过程这一零假设，因此可以认为时间序列来自的系统不具有非线性特征，与实际情况一致。

零假设 2：时间序列来自具常系数的高斯线性随机过程经过单调静态非线性变换。

利用 2.5.1 节中迭代的幅度调节 Fourier 变换（IAAFT）方法生产 19 组替代三变量时间序列，经计算，线性冗余的显著性检验统计量 $\mathrm{DIFF}L(x_{1,n};x_{2,n+\tau};x_{3,n})$ 关于时间延迟间隔 τ 的函数如图 3.14(a) 所示，广义冗余显著性检验统计量 $\mathrm{DIFF}R_q(x_{1,n};x_{2,n+\tau};x_{3,n})$ 关于时间延迟间隔 τ 的函数如图 3.14(b)($q=1$)、图 3.14(c)($q=2$) 和图 3.14(d)($q=3$) 所示。

线性边缘冗余 $\lambda(x_{1,n};x_{2,n+\tau};x_{3,n})$ 与广义边缘冗余 $\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n})$ 的显著性检验统计量 $\mathrm{DIFF}\lambda(x_{1,n};x_{2,n+\tau};x_{3,n})$ 和 $\mathrm{IDFF}\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n})$ 如图 3.15 所示。

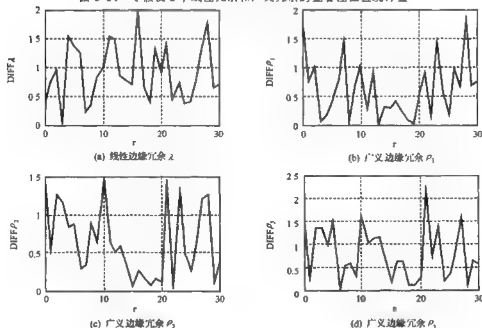
可见，线性冗余的显著性检验统计量都很小，这表明原时间序列与替代时间序



(a) 线性冗余 L　　(b) 广义冗余 R_1

(c) 广义冗余 R_2　　(d) 广义冗余 R_3

图 3.14　零假设 2 下线性冗余和广义冗余的显著性检验统计量



(a) 线性边缘冗余 λ　　(b) 广义边缘冗余 ρ_1

(c) 广义边缘冗余 ρ_2　　(d) 广义边缘冗余 ρ_3

图 3.15　零假设 2 下线性边缘冗余和广义边缘冗余的显著性检验统计量

列的线性依赖结构没明显差异,即生成的替代时间序列是好的。同时广义冗余的显著性检验统计量的值几乎都小于 1.96,即不能有效地拒绝时间序列来自具常系数的高斯线性随机过程经过静态非线性变换这一零假设,因此可以认为时间序列来自的系统不具有非线性特征,与实际情况一致。

第四章　多变量时间序列相空间重构

在实际问题中，一个复杂系统往往是由多个状态变量描述的，因此，通过观测或者实验可以从不同的角度获得多维的观测结果，这些离散的观测量在时间轴上形成一组多变量时间序列。例如对生理学中的心脏这个复杂动力系统，心电图是常用的观测数据，目前广泛采用 8 路或者 16 路探头形成多维输出，从不同的角度对心脏的动力系统进行观察。对于证券市场这个复杂动力系统，如果把每只股票作为一个观测量，则证券市场同一时刻产生的观测数据多达几百维。原则上，对自治系统，根据 F. Takens 的延迟嵌入定理，只要嵌入维数足够大，单变量时间序列足以重构原动力系统，但在实际问题中，情形不完全是这样。例如，式(1.6)的 Lorenz 系统中利用 x 坐标可以较好地重构原动力系统，而利用 z 坐标就不能很好地重构原动力系统[103]。因此，不能保证实际问题中任何获取的单变量时间序列都能较好地重构原动力系统，特别在具有噪声的观测时间序列中，利用多变量时间序列重构可以获得比单变量时间序列重构更好的结果。另外，对比较短的观测时间序列，利用多变量时间序列重构可以弥补时间序列长度不足的问题。

4.1　多变量时间序列相空间重构的流程

假设从复杂系统中通过观测或实验手段获得一组多变量时间序列，首先需解决的问题是这些变量是否都需要。显然，对某些复杂系统，可以从不同的角度和用不同的实验方法获得多个变量的时间序列，但某两个变量的时间序列之间可能是相互依赖的，或者说某两个变量的时间序列可能反映的是复杂系统的同一特征，那么这两个变量的时间序列只需考虑一个即可，否则会导致信息的冗余并增加计算的复杂程度。例如，把一个病人看成一个复杂系统，则通过测量病人的心电图，可以获得心电周期(即两个连续的 QRS 复合波中 R 波之间的距离)时间序列，也可获得心率变异(即心电周期的倒数)时间序列，还可以获得两个连续的 QRS 复合波中 R 波峰间波形曲线下的面积时间序列及 R 波峰间波形曲线长度时间序列等，显然这些时间序列都是反映病人的心脏异常情况的，研究其中一个时间序列即可。

从另一个角度看，复杂系统通常分解成几个子系统进行研究，假设从每一子系统通过观测或实验各获得一组单变量时间序列，组成整个复杂系统的一组多变量时间序列，要解决的问题是这些变量是否都需要，或者说这些子系统是否是完全独立的？如果一组单变量时间序列完全依赖于另一组单变量时间序列，即一组单变量

时间序列可由另一组单变量时间序列预测或它们之间存在一个映射关系，则这组单变量时间序列是多余的，在多变量时间序列的重构中，这样的时间序列可不考虑。

因此，当我们通过观测或实验手段获得实际问题中的复杂系统的多变量时间序列后，首先需要研究的问题是这些变量的时间序列之间是否有某种依赖关系，如果某一变量的时间序列依赖于另外一些变量的时间序列，则在研究相空间重构时可以去掉，这样逐一处理后可以得到一组相互独立的多变量时间序列。

多变量时间序列相空间重构最早的重构方式比较简单，就是每一变量作为重构的多维空间中的一个分量，重构空间的维数与变量个数相等，与延迟时间无关，这样做的好处就是避免了重构参数的选取问题，但当测量的变量个数小于原动力系统的实际维数时，这样做显然会失去某些信息，重构的空间不等价于原动力系统。为了能从多变量时间序列通过延迟嵌入方法重构相空间，需要讨论延迟时间间隔和嵌入维数的确定方法，但与单变量时间序列相比，重构参数的确定会更困难，因为这时延迟时间间隔和嵌入维数分别有多个，而不是各一个。

相空间的嵌入维数是系统复杂性的一个界，这种拓扑维数一般情况下只是吸引子占据空间数量的一个非常粗糙的描述，为了给出更严格的描述，有不少不同的维数概念，其中关联维数是最常用的一种维数度量，尤其对观测或实验获得的时间序列，可由 G P 算法直接计算。关联维数度量了研究系统的复杂性，刻画了相空间中点的分布，在确定性系统中，关联维数就是生成相应系统所必须的独立变量的个数，规则的确定性系统有整数关联维数，而混沌系统有非整数关联维数，大于此关联维数的下一个整数就是系统的独立变量的个数。如何从多变量时间序列重构的相空间中计算关联维数是需要进行研究的问题。

在动力系统分析中，Lyapunov 指数是一个十分重要的概念，一个非线性系统的全部 Lyapunov 指数称为该系统的 Lyapunov 指数谱，它与初始状态无关，在拓扑映射下不变，是复杂系统刻画和分类的重要指标。最大 Lyapunov 指数定量地刻画了初始靠近的状态空间轨线的指数发散率，特别最大 Lyapunov 指数为正是混沌的重要特征。因此如何在多变量时间序列中有效地计算 Lyapunov 指数也是需要研究的问题。

综上所述，多变量时间序列相空间重构的流程可以用图 4.1 表示。

多变量观测时间序列中往往包含比单变量观测时间序列更丰富的关于观测系统的信息，因此，对多变量时间序列进行相空间重构可以更好地获取观测系统的特征和演化行为，通过在重构的相空间中计算几何不变量，可以分析、解释、指导实际系统，因此，这时有许多问题需要进一步进行研究。

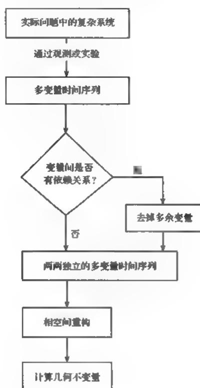


图 4.1　多变量时间序列相空间重构的流程图

4.2　多变量时间序列中变量间的依赖关系

通过观测或实验获得的多变量时间序列中各观测变量之间可能存在一定的相关性，如何剔除其中的冗余数据，寻找能够反映动力系统演化特征的数据是首先需要解决的一个问题。

对实际问题中的复杂系统，可以根据问题的实际意义对获得的 K 个变量的时间序列进行分类，分成 L 个子集，下面主要研究子集 I_l 中的一个时间序列是否依赖于其余 K_l-1 个时间序列这一问题。

假设 $\{x_n\}_{n=1}^{N}$ 和 $\{y_n\}_{n=1}^{N}$ 分别是某一复杂系统的两个子系统 X 和 Y 通过观测或实验产生的时间序列，如果它们在统计上是依赖的，称子系统 X 和 Y 是耦合的，这意味着可以从子系统 Y 中获取子系统 X 的某些信息，反之亦然。测量两个平稳时间序列间统计依赖性的最基本的方法是正交关联函数法，这种方法的优点是计算简单方便，但它只能测量线性依赖性，无法测量非线性依赖性，虽然也可用高阶正

交函数测量非线性依赖性，但计算变得复杂多了。在非线性时间序列分析中，互信息法是测量统计依赖性的一种常用方法，当 X 和 Y 耦合时，它们之间有公共的信息，这种公共信息的多少可以由 $\{x_n\}_{n=1}^{N}$ 和 $\{y_n\}_{n=1}^{N}$ 之间的互信息来度量。当子系统 X 和 Y 没有耦合时，意味着 X 和 Y 之间没有公共信息，即互信息为零，此时从统计意义看从给定的 $\{y_n\}_{n=1}^{N}$ 不能获得 $\{x_n\}_{n=1}^{N}$ 的信息，反之亦然，表明此时获得的这两个时间序列是独立的；当子系统 X 和 Y 有最大耦合时，意味着两个子系统之间有确定性关系，此时，几乎对所有的 n，$\{y_n\}_{n=1}^{N}$ 是 $\{x_n\}_{n=1}^{N}$ 的函数，反之亦然，在这种情况下，表明其中一个时间序列是多余的，因此，在整个复杂系统的研究中，当两个子系统具最大耦合时，只需考虑其中一个即可。由此看来，互信息可以认为是两个系统耦合强度的一种度量，能较好地反映两个系统的耦合情况，但它的计算比较麻烦，需用直方图的方法求分布密度函数，为了克服这一困难，文献[104,105]提出用广义互信息测量统计依赖性的方法，但要求其中一个时间序列满足一致分布，否则要通过复杂的转换才能使用。我们对文献[104,105]中的方法做了改进和推广，提出了一种新的统计依赖性测量方法，这种方法不需要预先假设时间序列满足一致分布[106,107]。

4.2.1　随机变量间统计依赖性的度量方法

设 ξ,η 分别是密度函数为 $p_1(x)$，$p_2(y)$ 的两个随机变量，(ξ,η) 的联合密度函数为 $p(x,y)$，定义 ξ,η 的正交关联函数为

$$\mathrm{Cor}(\xi,\eta)=\frac{\overline{\xi\eta}-\overline{\xi}\,\overline{\eta}}{\sigma_\xi\sigma_\eta} \tag{4.1}$$

式中，$\overline{\xi\eta}$，$\overline{\xi}$，$\overline{\eta}$ 分别是随机变量 (ξ,η)，ξ,η 的均值；σ_ξ,σ_η 分别是随机变量 ξ,η 的标准差。$\mathrm{Cor}(\xi,\eta)=0$ 的充分必要条件是 ξ 与 η 在统计上不关联。定义 ξ,η 的互信息为

$$\begin{aligned}I_1(\xi,\eta)&=H_1(\xi)+H_1(\eta)-H_1(\xi,\eta)\\&=\iint p(x,y)\log_2\frac{p(x,y)}{p_1(x)p_2(y)}\mathrm{d}x\mathrm{d}y\end{aligned} \tag{4.2}$$

其中

$$H_1(\xi)=-\int p_1(x)\log_2 p_1(x)\mathrm{d}x \tag{4.3}$$

$$H_1(\eta)=-\int p_2(y)\log_2 p_2(y)\mathrm{d}y \tag{4.4}$$

$$H_1(\xi,\eta)=-\iint p(x,y)\log_2 p(x,y)\mathrm{d}x\mathrm{d}y \tag{4.5}$$

分别是 (ξ,η)，ξ,η 的 Shannon 熵。不难证明：

$$0\leqslant I_1(\xi,\eta)\leqslant\min\{H_1(\xi),H_1(\eta)\} \tag{4.6}$$

且 $I_1(\xi,\eta)=0$ 的充分必要条件是 ξ,η 在统计上独立，即 $p(x,y)=p_1(x)p_2(y)$；

$I_1(\xi,\eta)=H_1(\xi)$[或 $H_1(\eta)$]的充分必要条件是 η 统计上依赖于 ξ(或 ξ 统计上依赖于 η)。

对观测或实验产生的时间序列 $\{x_n\}_{n=1}^N$ 和 $\{y_n\}_{n=1}^N$，这种方法的困难在于需要利用直方图计算其密度函数，计算过程比较麻烦。为了解决这一问题，文献[104，105]提出了用 Renyi 熵代替 Shannon 熵的广义互信息：

$$I_2(\xi,\eta)=H_2(\xi)+H_2(\eta)-H_2(\xi,\eta)$$

$$=\log_2\frac{\iint p^2(x,y)\mathrm{d}x\mathrm{d}y}{\int p_1^2(x)\mathrm{d}x\int p_2^2(y)\mathrm{d}y} \tag{4.7}$$

其中

$$H_2(\xi)=-\log_2\int p_1^2(x)\mathrm{d}x \tag{4.8}$$

$$H_2(\eta)=-\log_2\int p_2^2(y)\mathrm{d}y \tag{4.9}$$

$$H_2(\xi,\eta)=-\log_2\iint p^2(x,y)\mathrm{d}x\mathrm{d}y \tag{4.10}$$

分别是 (ξ,η)，ξ，η 的 Renyi 熵。

这一方法的优点是对观测或实验产生的时间序列，计算 Renyi 熵比较简单，可用关联积分来计算，但问题在于 $I_2(\xi,\eta)$ 可能是负的，且 $I_2(\xi,\eta)=0$ 不一定意味着 ξ，η 在统计上独立。为了克服这一困难，文献[104，105]中要求 ξ，η 中有一随机变量服从一致分布。例如，η 的密度函数为 $p_2(y)\equiv\varepsilon$，则 $I_2(\xi,\eta)$ 成为

$$I_2(\xi,\eta)=\log_2\frac{\iint p^2(x,y)\mathrm{d}x\mathrm{d}y}{\varepsilon\int p_1^2(x)\mathrm{d}x} \tag{4.11}$$

此时

$$0\leqslant I_2(\xi,\eta)\leqslant\min\{H_2(\xi),H_2(\eta)\} \tag{4.12}$$

且 $I_2(\xi,\eta)=0$ 的充分必要条件是 ξ，η 在统计上独立，$I_2(\xi,\eta)=H_2(\xi)$[或 $H_2(\eta)$]的充分必要条件是 η 统计上依赖于 ξ(或 ξ 统计上依赖于 η)[105]。当 ξ，η 不服从一致分布时，文献[105]说明了可通过对原始时间序列作相关变换，使它成为一致分布，并证明了此时广义互信息的有关性质不变，但这样一来又带来了计算上的困难。

为了能对任何随机变量 ξ，η 直接应用，我们定义一种新的统计依赖性度量：

$$I_2^*(\xi,\eta)=\log_2\left[\frac{\iint[p(x,y)-p_1(x)p_2(y)]^2\mathrm{d}x\mathrm{d}y}{\int p_1^2(x)\mathrm{d}x\int p_2^2(y)\mathrm{d}y}+1\right]$$

$$-\log_2\left[\frac{\iint p^2(x,y)\mathrm{d}x\mathrm{d}y}{\int p_1^2(x)\mathrm{d}x\int p_2^2(y)\mathrm{d}y}+2\left(1-\frac{\iint p(x,y)p_1(x)p_2(y)\mathrm{d}x\mathrm{d}y}{\int p_1^2(x)\mathrm{d}x\int p_2^2(y)\mathrm{d}y}\right)\right] \tag{4.13}$$

显然 $I_2^*(\xi,\eta)\geqslant 0$，且 $I_2^*(\xi,\eta)=0$ 的充分必要条件是 ξ 与 η 在统计上独立。由于

$$\begin{aligned}\iint p^2(x,y)\mathrm{d}x\mathrm{d}y&=\int\left(\int p^2(x,y)\mathrm{d}y\right)\mathrm{d}x\\&\leqslant\int\left(\int p(x,y)\mathrm{d}y\right)^2\mathrm{d}x=\int p_1^2(x)\mathrm{d}x\end{aligned} \tag{4.14}$$

则

$$I_2^*(\xi,\eta)\leqslant\log_2\left[2+\frac{1}{\int p_2^2(y)\mathrm{d}y}\left(1-\frac{2\iint p(x,y)p_1(x)p_2(y)\mathrm{d}x\mathrm{d}y}{\int p_1^2(x)\mathrm{d}x}\right)\right] \tag{4.15}$$

若记 $p_{2,\min}=\min\{p_2(y)\mid y\in\eta\}$，$p_{2,\max}=\max\{p_2(y)\mid y\in\eta\}$，则

$$p_{2,\min}\int p_1^2(x)\mathrm{d}x\leqslant\iint p(x,y)p_1(x)p_2(y)\mathrm{d}x\mathrm{d}y\leqslant p_{2,\max}\int p_1^2(x)\mathrm{d}x \tag{4.16}$$

因此可得 $I_2^*(\xi,\eta)$ 的最大值 $I_{2,\max}^*(\xi,\eta)$ 的一个近似估计，即

$$\log_2\left(2+\frac{1-2p_{2,\max}}{\int p_2^2(y)\mathrm{d}y}\right)\leqslant I_{2,\max}^*(\xi,\eta)\leqslant\log_2\left(2+\frac{1-2p_{2,\min}}{\int p_2^2(y)\mathrm{d}y}\right) \tag{4.17}$$

当 $p_{2,\max}\gg p_{2,\min}$，且 $p_{2,\max}\approx\frac{1}{2}$ 时，上述估计没有实际用处，典型的情况是 $p_{2,\max}\ll 1$，甚至 $p_{2,\max}\to 0$。当观察时间序列足够长，且 $p_{2,\max}\to 0$ 时，$\int p_2^2(y)\mathrm{d}y\to 0$，因此

$$I_{2,\max}^*(\xi,\eta)\approx\log_2\left(2+\frac{1}{\int p_2^2(y)\mathrm{d}y}\right)\approx\log_2\left(\frac{1}{\int p_2^2(y)\mathrm{d}y}\right) \tag{4.18}$$

同理，当 $p_{1,\max}\to 0$ 时

$$I_{2,\max}^*(\xi,\eta)\approx\log_2\left(\frac{1}{\int p_2^2(y)\mathrm{d}y}\right) \tag{4.19}$$

另外，当 η 在统计上依赖于 ξ 时，对某个 $y=y(x)$，$p(x,y)=p_1(x)$，对其他 $y\neq y(x)$，$p(x,y)\equiv 0$，因此

$$I_2^*(\xi,\eta)=\log_2\left[2+\frac{1}{\int p_2^2(y)\mathrm{d}y}-\frac{2\iint p_1^2(x)p_2(y(x))\mathrm{d}x}{\int p_1^2(x)\mathrm{d}x\int p_2^2(y)\mathrm{d}y}\right] \tag{4.20}$$

由于 $p_{2,\min}\leqslant p_2(y(x))\leqslant p_{2,\max}$，此时 $I_2^*(\xi,\eta)$ 在 $\log_2\left(2+\frac{1-2p_{2,\max}}{\int p_2^2(y)\mathrm{d}y}\right)$ 与

$\log_2\left[2+\frac{1-2p_{2,\min}}{\int p_2^2(y)\mathrm{d}y}\right]$之间，表明$I_2^*(\xi,\eta)$近似取到最大值$I_{2,\max}^*(\xi,\eta)$。当$\xi$在统计上依赖于$\eta$时，可同理讨论。

综上可得：$0\leqslant I_2^*(\xi,\eta)\leqslant I_{2,\max}^*(\xi,\eta)$，而且$I_2^*(\xi,\eta)=0$的充分必要条件是$\xi$与$\eta$在统计上独立；当$\eta$在统计上依赖于$\xi$(或$\xi$在统计上依赖于$\eta$)时，$I_2^*(\xi,\eta)$近似取到其最大值$I_{2,\max}^*(\xi,\eta)$，其中$I_{2,\max}^*(\xi,\eta)$由式(4.18)[或式(4.19)]近似估计。

注：当η服从一致分布时，即$p_2(y)\equiv\varepsilon$时

$$I_2^*(\xi,\eta)=\log_2\frac{\iint p^2(x,y)\mathrm{d}x\mathrm{d}y}{\varepsilon\int p_1^2(x)\mathrm{d}x} \tag{4.21}$$

成为文献[104,105]中定义的广义互信息。

4.2.2　观测时间序列统计依赖性的计算

1. 两组时间序列间的统计依赖性

设$\{x_n\}_{n=1}^N$和$\{y_n\}_{n=1}^N$分别是子系统X和Y通过观测或实验获得的时间序列，按照 Takens 的延迟嵌入定理，分别做以下相空间重构

$$\boldsymbol{x}_n=(x_n,x_{n-\tau_x},\cdots,x_{n-(m_x-1)\tau_x}),\quad n=N_x,N_x+1,\cdots,N$$
$$N_x=(m_x-1)\tau_x+1 \tag{4.22}$$

$$\boldsymbol{y}_n=(y_n,y_{n-\tau_y},\cdots,y_{n-(m_y-1)\tau_y}),\quad n=N_y,N_y+1,\cdots,N$$
$$N_y=(m_y-1)\tau_y+1 \tag{4.23}$$

$$\boldsymbol{z}_n=(x_n,x_{n-\tau_x},\cdots,x_{n-(m_x-1)\tau_x},y_n,y_{n-\tau_y},\cdots,y_{n-(m_y-1)\tau_y}),\quad n=N_z,N_z+1,\cdots,N$$
$$N_z=\max\{N_x,N_y\} \tag{4.24}$$

其中正整数$\tau_x(\tau_y)$和$m_x(m_y)$分别是延迟时间间隔和嵌入维数。以最大范数$\|\cdot\|_{\max}$定义两点间的距离，以下为简单，记为$\|\cdot\|$。按照 Takens 的延迟嵌入定理，确定了这些参数后，状态空间$\boldsymbol{x}_n\to\boldsymbol{x}_{n+1}$(或$\boldsymbol{y}_n\to\boldsymbol{y}_{n+1}$)的演化反映了子系统$X$(或$Y$)的演化，并且相空间中吸引子的几何特征与原子系统$X$(或$Y$)中的几何特征等价。

设$\boldsymbol{x}_i$和$\boldsymbol{y}_j$分别是随机向量$\xi=(\xi_1,\xi_2,\cdots,\xi_{m_x})$和$\eta=(\eta_1,\eta_2,\cdots,\eta_{m_y})$的实现，记

$$p(x,y)=\mathrm{prob}\{\xi_1=x_i,\cdots,\xi_{m_x}=x_{i+(m_x-1)\tau_x},\eta_1=y_j,\cdots,\eta_{m_y}=y_{j+(m_y-1)\tau_y}\} \tag{4.25}$$

$$p_1(x)=\sum_y p(x,y)=\mathrm{prob}\{\xi_1=x_i,\cdots,\xi_{m_x}=x_{i+(m_x-1)\tau_x}\} \tag{4.26}$$

$$p_2(y)=\sum_x p(x,y)=\mathrm{prob}\{\eta_1=y_j,\cdots,\eta_{m_y}=y_{j+(m_y-1)\tau_y}\} \tag{4.27}$$

对 z_n 做 ε 剖分 β_ε，分成一系列以 2ε 为边长的超立方体。考虑点 z_i，记 $B_{m(i)}$ 为 β_ε 中包含 z_i 的超立方体。假设 ε 充分小，点 z_i 在 $B_{m(i)}$ 中的概率 $p_{m(i)}$ 在超立方体被改变为 z_i 的 ε 邻域时不会发生本质的改变，则此概率在时间序列具遍历性的假设下可由时间的平均估计，即

$$p_{m(i)} \approx \lim_{n\to\infty} \frac{1}{N-N_z} \sum_{j=N_z,j\neq i}^{N} H(\varepsilon - \| z_i - z_j \|) \tag{4.28}$$

其中 $H(x)=\begin{cases}0, x\leqslant 0\\ 1, x>0\end{cases}$ 是 Heavside 函数。因此，当 $N\to\infty$，$\varepsilon\to 0$ 时，有

$$\begin{aligned}\iint p^2(x,y)\mathrm{d}x\mathrm{d}y &\approx \sum_x\sum_y p^2(x,y) \approx \sum_i p_i^2 = \frac{1}{N-N_z+1}\sum_{i=N_z}^{N} p_{m(i)} \\ &\approx \frac{1}{(N-N_z)(N-N_z+1)}\sum_{i\neq j} H(\varepsilon - \| z_i - z_j \|) \\ &= C_2(Z,\varepsilon)\end{aligned} \tag{4.29}$$

其中 p_i 是 (x_i,y_i) 在第 i 个超立方体内的概率，$C_2(Z,\varepsilon)$ 是按式(4.24)重构的相空间中的关联积分。同理可得

$$\begin{aligned}\int p_1^2(x)\mathrm{d}x &\approx \sum_x p_1^2(x) \approx \frac{1}{(N-N_x)(N-N_x+1)}\sum_{i\neq j} H(\varepsilon - \| x_i - x_j \|) \\ &= C_2(X,\varepsilon)\end{aligned} \tag{4.30}$$

$$\begin{aligned}\int p_2^2(y)\mathrm{d}y &\approx \sum_y p_2^2(y) \approx \frac{1}{(N-N_y)(N-N_y+1)\sum_{i\neq j} H(\varepsilon - \| y_i - y_j \|)} \\ &= C_2(Y,\varepsilon)\end{aligned} \tag{4.31}$$

进一步，若对 x_n，y_n 分别作 ε 剖分 $\beta_{1,\varepsilon}=\{B_{1,m}\}$，$\beta_{2,\varepsilon}=\{B_{2,n}\}$，分成一系列以 2ε 为边长的超立方体。考虑点 z_i，它的投影 $x_i\in B_{1,m(i)}$，$y_i\in B_{2,m(i)}$，则点 $x_i\in B_{1,m(i)}$ 的概率为

$$p_{1,m(i)} \approx \lim_{N\to\infty} \frac{1}{N-N_x} \sum_{j=N_x,j\neq i}^{N} H(\varepsilon - \| x_i - x_j \|) \tag{4.32}$$

点 $y_i\in B_{2,n(i)}$ 的概率为

$$p_{2,m(i)} \approx \lim_{N\to\infty} \frac{1}{N-N_y} \sum_{j=N_y,k\neq i}^{N} H(\varepsilon - \| y_i - y_k \|) \tag{4.33}$$

因此，当 $N\to\infty$，$\varepsilon\to 0$ 时，有

$$\begin{aligned}\iint p(x,y)p_1(x)p_2(y)\mathrm{d}x\mathrm{d}y &\approx \sum_x\sum_y p(x,y)p_1(x)p_2(y) \\ &\approx \sum_i p_i p_{1,i} p_{2,i} = \frac{1}{N-N_z+1}\sum_{i=N_z}^{N} p_{1,m(i)}p_{2,m(i)}\end{aligned}$$

$$\approx \frac{1}{(N-N_z+1)(N-N_x)(N-N_y)}$$
$$\times \sum_{i=N_z}^{N} \sum_{j=N_x, j\neq i}^{N} \sum_{k=N_y, k\neq i}^{N} H(\varepsilon - \| x_i - x_j \|) H(\varepsilon - \| y_i - y_k \|)$$
$$= K_2(X,Y,Z,\varepsilon) \tag{4.34}$$

式中，$p_t, p_{1,t}, p_{2,t}$ 分别是 (x_i, y_j)，x_i，y_j 在第 t 个超立方体 $B_{1,m}\times B_{2,n}$，$B_{1,m}$，$B_{2,n}$ 内的概率。

由此可得 4.2.1 节定义的统计依赖性度量及其最大值的关联积分的估计，即

$$I_2^*(X,Y) = I_2^*(\xi,\eta)$$
$$= \log_2\left[\frac{C_2(Z,\varepsilon)}{C_2(X,\varepsilon)C_2(Y,\varepsilon)} + 2\left(1 - \frac{K_2(X,Y,Z,\varepsilon)}{C_2(X,\varepsilon)C_2(Y,\varepsilon)}\right)\right] \tag{4.35}$$

$$I_{2,\max}^{*,(1)}(X,Y) = I_{2,\max}^{*,(1)}(\xi,\eta) \approx \log_2\left(\frac{1}{C_2(X,\varepsilon)}\right) \tag{4.36}$$

$$I_{2,\max}^{*,(2)}(X,Y) = I_{2,\max}^{*,(2)}(\xi,\eta) \approx \log_2\left(\frac{1}{C_2(Y,\varepsilon)}\right) \tag{4.37}$$

2. 多组时间序列间的统计依赖性

设 $\{x_{l,n}\}_{n=1}^{N}(l=1,2,\cdots,L)$ 分别是子系统 $X_l(l=1,2,\cdots,L)$ 通过观测或实验获得的时间序列，任何两组时间序列间的统计依赖性可由上面的方法计算。实际问题中有可能出现以下情况，即任何两组时间序列在统计上是独立的，但其中一组时间序列（如 $\{x_{L,n}\}_{n=1}^{N}$）依赖于其余 $L-1$ 组时间序列［如 $\{x_{l,n}\}_{n=1}^{N}(l=1,2,\cdots,L-1)$］，为了计算这种统计依赖性，分别做以下相空间重构

$$x_{L,n} = (x_{L,n}, x_{L,n-\tau_L}, \cdots, x_{L,n-(m_L-1)\tau_L}),\ n = N_L, N_L+1, \cdots, N$$
$$N_L = (m_L - 1)\tau_L + 1 \tag{4.38}$$

$$z_{L-1,n} = (x_{1,n}, x_{1,n-\tau_1}, \cdots, x_{1,n-(m_1-1)\tau_1}, x_{2,n}, x_{2,n-\tau_2}, \cdots, x_{2,n-(m_2-1)\tau_2}, \cdots,$$
$$x_{2,n-(m_2-1)\tau_2}, \cdots, x_{L-1,n}, x_{L-1,n-\tau_{L-1}}, \cdots, x_{L-1,n-(m_{L-1}-1)\tau_{L-1}})$$
$$n = N_L^{(1)}, N_L^{(1)}+1, \cdots, N;\quad N_L^{(1)} = \max_{1\leqslant l\leqslant L-1}(m_l-1)\tau_l + 1 \tag{4.39}$$

$$z_{L,n} = (x_{1,n}, x_{1,n-\tau_1}, \cdots, x_{1,n-(m_1-1)\tau_1}, x_{2,n}, x_{2,n-\tau_2}, \cdots, x_{2,n-(m_2-1)\tau_2}, \cdots,$$
$$x_{2,n-(m_2-1)\tau_2}, \cdots, x_{L,n}, x_{L,n-\tau_L}, \cdots, x_{L,n-(m_L-1)\tau_L})$$
$$n = N_L^{(2)}, N_L^{(2)}+1, \cdots, N;\ N_L^{(2)} = \max_{1\leqslant l\leqslant L}(m_l-1)\tau_l + 1 \tag{4.40}$$

则类似于前面的讨论，可得时间序列 $\{x_{L,n}\}_{n=1}^{N}$ 与其余 $L-1$ 组时间序列 $\{x_{l,n}\}_{n=1}^{N}(l=1,2,\cdots,L-1)$ 的统计依赖性度量及其最大值的关联积分的估计，即

$$I_2^*(X_1,\cdots,X_{L-1};X_L) \approx \log_2\left[\frac{C_2(Z_L,\varepsilon)}{C_2(X_L,\varepsilon)C_2(Z_{L-1},\varepsilon)} + 2\left(1 - \frac{K_2(L_L,Z_{L-1},Z_L,\varepsilon)}{C_2(X_L,\varepsilon)C_2(Z_{L-1},\varepsilon)}\right)\right] \tag{4.41}$$

$$I_{2,\max}^{*,(1)}(X_1,\cdots,X_{L-1};X_L) \approx \log_2\left(\frac{1}{C_2(X_L,\varepsilon)}\right) \tag{4.42}$$

$$I_{2,\max}^{*,(2)}(X_1,\cdots,X_{L-1};X_L) \approx \log_2\left(\frac{1}{C_2(Z_{L-1},\varepsilon)}\right) \tag{4.43}$$

其中

$$C_2(X_L,\varepsilon) = \frac{1}{(N-N_L)(N-N_L+1)}\sum_{i\neq j} H(\varepsilon - \| x_{L,i} - x_{L,j} \|) \tag{4.44}$$

$$C_2(Z_{L-1},\varepsilon) = \frac{1}{(N-N_L^{(1)})(N-N_L^{(1)}+1)}\sum_{i\neq j} H(\varepsilon - \| z_{L-1,i} - z_{L-1,j} \|) \tag{4.45}$$

$$C_2(Z_L,\varepsilon) = \frac{1}{(N-N_L^{(2)})(N-N_L^{(2)}+1)}\sum_{i\neq j} H(\varepsilon - \| z_{L,i} - z_{L,j} \|) \tag{4.46}$$

$$K_2(X,Z_{L-1},Z_L,\varepsilon) = \frac{1}{(N-N_L^{(2)}+1)(N-N_L)(N-N_L^{(1)})}$$
$$\times \sum_{i=N_L^{(2)}}^{N} \sum_{j=N_L, j\neq i}^{N} \sum_{k=N_L^{(1)}, k\neq i}^{N} H(\varepsilon - \| x_{L,i} - x_{L,j} \|) H(\varepsilon - \| z_{L-1,i} - z_{L-1,k} \|) \tag{4.47}$$

4.2.3　应用举例

例 4.1　设随机向量(ξ,η)的联合分布密度为$p_{11} = -\frac{1}{2} + 2p - p^2$，$p_{22} = \frac{1}{2} - p^2$，$p_{12} = p_{21} = \frac{1}{2} - p + p^2$，其中$1 - \frac{1}{\sqrt{2}} < p < \frac{1}{\sqrt{2}}$，因此$\xi,\eta$有相同的概率分布$p_1 = p_{11} + p_{12} = p$，$p_2 = p_{21} + p_{22} = 1 - p$，则

$$I_1(\xi,\eta) = \log_2\left[\left(\frac{1}{2} - p^2\right)^{\frac{1}{2}-p^2}(1-p)^{2p-2}p^{-2p}\left(\frac{1}{2} - p + p^2\right)^{1-2p+2p^2} \times \left(-\frac{1}{2} + 2p - p^2\right)^{-\frac{1}{2}+2p-p^2}\right] \tag{4.48}$$

$$I_2(\xi,\eta) = 0 \tag{4.49}$$

$$I_2^*(\xi,\eta) = \log_2 \frac{2 - 12p + 32p^2 - 40p^3 + 20p^4}{1 - 4p + 8p^2 - 8p^3 + 4p^4} \tag{4.50}$$

由于只有在$p = \frac{1}{2}$时，$p_{ij} = p_i p_j$，$i,j = 1,2$，即ξ与η在统计上独立，因此广义互信息无法检测到$p \neq \frac{1}{2}$时ξ与η的统计依赖性，而$I_1(\xi,\eta) = 0$，$I_2^*(\xi,\eta) = 0$的充

分必要条件是 $p=\frac{1}{2}$，因此互信息和 4.2.1 节定义的统计依赖性度量能很好地检测到 $p\neq\frac{1}{2}$ 时 ξ 与 η 的统计依赖性。

例 4.2　设随机向量 (ξ,η) 的联合分布密度函数为

$$p_\delta(x,y)=\begin{cases}1+\delta,(x,y)\in([0,0.5]\times[0,0.5]\cup([0.5,1]\times[0.5,1])\\1-\delta,(x,y)\in([0,0.5]\times[0.5,1])\cup([0.5,1]\times[0,0.5])\\0,(x,y)\in \text{其他}\end{cases}$$

(4.51)

则 $p_\delta(x)=p_\delta(y)=1$，$\xi,\eta$ 是 $[0,1]$ 上的一致分布，其平均值为 $\bar{\xi}=\bar{\eta}=\frac{1}{2}$，标准差为 $\sigma_\xi=\sigma_\eta=\frac{1}{\sqrt{12}}$，$\overline{\xi\eta}=\frac{1}{4}+\frac{1}{16}\delta$，因此正交关联函数为

$$\mathrm{Cor}(\xi,\eta)=\frac{3}{4}\delta \tag{4.52}$$

计算可得

$$I_1(\xi,\eta)=\frac{(1+\delta)\log_2(1+\delta)-(1-\delta)\log_2(1-\delta)}{2} \tag{4.53}$$

$$I_2(\xi,\eta)=\log_2(1+\delta^2) \tag{4.54}$$

$$I_2^*(\xi,\eta)=\log_2(1+\delta^2) \tag{4.55}$$

当 $\delta=1$ 时，$\mathrm{Cor}(\xi,\eta)=\frac{3}{4}\delta$，$I_1(\delta,\eta)=I_2(\xi,\eta)=I_2^*(\xi,\eta)=1$ 都取到最大值，由于 $\delta=\frac{4}{3}\mathrm{Cor}(\xi,\eta)$，因此 $I_1(\xi,\eta)$，$I_2(\xi,\eta)$，$I_2^*(\xi,\eta)$ 都是 $\mathrm{Cor}(\xi,\eta)$ 的函数。在这个特殊的例子中，ξ,η 在统计上独立的充分必要条件是 ξ,η 不关联。但由于任何两个随机变量的互关联函数不会超过 1，因此 $\frac{3}{4}$ 是一个相当大的值，表明具有较强的关联性，而对 $I_1(\xi,\eta)$，$I_2(\xi,\eta)$，$I_2^*(\xi,\eta)$，它们的取值在 0 到 $+\infty$ 之间，1 是一个非常小的值，表明此时非线性统计依赖性不强。这个例子说明了 ξ 与 η 关联性较强，即线性依赖性较强，但非线性依赖性较弱。

例 4.3　考虑下面的耦合 Henon 映射

$$X:\begin{cases}x_{n+1}=1-1.4x_n^2+0.3u_n\\u_{n+1}=x_n\end{cases} \tag{4.56}$$

$$Y:\begin{cases}y_{n+1}=1-1.4[Cx_n+(1-C)y_n]y_n+0.3v_n\\v_{n+1}=y_n\end{cases} \tag{4.57}$$

其中耦合强度 C 在 0～1 之间。取初始点 $x_0=u_0=0$，$y_0=v_0=1$，迭代得 2024 个点，去掉前 1000 个点，获时间序列 $\{x_n\}_{n=1}^{1024}$ 和 $\{y_n\}_{n=1}^{1024}$。

图 4.2～图 4.4 给出了耦合强度 C 分别为 0.1，0.65，0.9 时，x_i-y_i 关系图，从

图 4.2～图 4.4 中可以看出，当 $C=0.1$ 时，几乎没有耦合，而当 $C=0.9$ 时，产生了几乎完美的耦合。

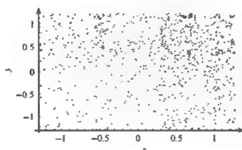


图 4.2　$C=0.1$ 时 x_i-y_i 关系图

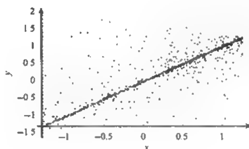


图 4.3　$C=0.65$ 时 x_i-y_i 关系图

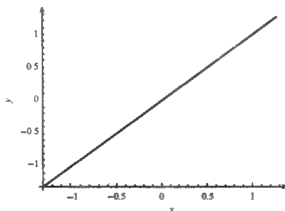


图 4.4　$C=0.9$ 时 x_i-y_i 关系图

图 4.5、图 4.6 和图 4.7 分别给出了 $C=0.1, 0.65, 0.9$ 时，用式(4.35)、

式(4.36)和式(4.37)计算得到的统计依赖性度量 $I_2^*(X,Y)$ 及它的最大值 $I_{2,\max}^{*,(1)}(X,Y)$ 或 $I_{2,\max}^{*,(2)}(X,Y)$，其中 $m_x=m_y=3$，$\tau_x=\tau_y=1$，ϵ 分别取 2^{-k}，$k=0,1,\cdots,10$。

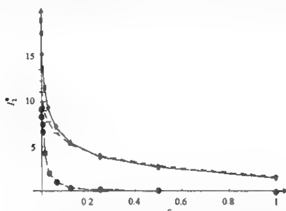


图 4.5　$C=0.1$ 时 I_2^*-ϵ 关系图，从上到下的曲线分别为 $I_{2,\max}^{*,(1)}(X,Y)$、$I_{2,\max}^{*,(2)}(X,Y)$ 和 $I_2^*(X,Y)$

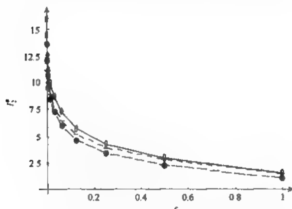


图 4.6　$C=0.65$ 时 I_2^*-ϵ 关系图，从上到下的曲线分别为 $I_{2,\max}^{*,(1)}(X,Y)$、$I_{2,\max}^{*,(2)}(X,Y)$ 和 $I_2^*(X,Y)$

从图 4.5 可知，只要 ϵ 不是太小，$I_2^*(X,Y)$ 几乎为零，且 $I_2^*(X,Y)$ 明显小于 $I_{2,\max}^*(X,Y)$，表明此时 $\{x_n\}_{n=1}^{1024}$ 和 $\{y_n\}_{n=1}^{1024}$ 统计上几乎独立，即系统 X 和 Y 几乎不耦合。从图 4.7 可知，$I_2^*(X,Y)$ 几乎等于 $I_{2,\max}^*(X,Y)$，表明此时 $\{x_n\}_{n=1}^{1024}$ 和 $\{y_n\}_{n=1}^{1024}$ 统计上相互依赖，即 X 和 Y 几乎耦合。图 4.6 的情况介于图 4.5 和图 4.7 的情况之间，$I_2^*(X,Y)$ 小于 $I_{2,\max}^*(X,Y)$，但 $I_2^*(X,Y)$ 对任何 ϵ 都不为零，表明此时 $\{x_n\}_{n=1}^{1024}$ 和 $\{y_n\}_{n=1}^{1024}$ 统计上不独立，但它们的统计依赖性也不是最强，即 X 和 Y 是耦合的，但耦合强度不是最大。以上结果与图 4.2、图 4.3 和图 4.4 的情况完全一致，表明提出

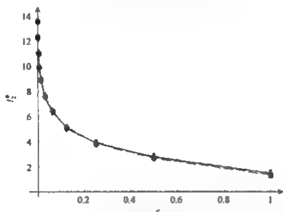


图 4.7　$C=0.9$ 时 I_2^c-ε 关系图，从上到下的曲线分别为 $I_{2,\max}^{c,(1)}(X,Y)$、$I_{2,\max}^{c,(2)}(X,Y)$ 和 $I_2^c(X,Y)$

的统计依赖性度量方法对这一例子是完全有效的。

4.3　多变量时间序列相空间重构参数的确定

假设从复杂系统中获得 K 个变量的时间序列，我们把它分成 L 个子集，即 I_1，I_2，…，I_L，子集 I_l 中含有 K_l 个变量的时间序列，$l=1,2,\cdots,L$，则 $K_1+K_2+\cdots+K_L=K$。如果子集 I_l 中的一个时间序列依赖于其余 K_l-1 个时间序列，则从 I_l 中只需取一个时间序列即可，这样可构成 $L(L<K)$ 个变量的两两独立的时间序列。

下面考虑有 L 个变量的两两独立的时间序列 $\{x_{l,n}\}_{n=1}^{N}$ $(l=1,2,\cdots,L)$，它是 L 个连续变量 $x_l(t)$ $(l=1,2,\cdots,L)$ 的测量值，即 $x_{l,n}=x_l(t_0+n\Delta t)$ $(n=1,2,\cdots,N;L=1,2,\cdots,L)$，其中 t_0 是初始时间，Δt 是样本时间。

早期对多变量时间序列的研究，都是在最简单的重构方式下讨论的，即每一变量作为重构的多维空间中的一个分量，重构空间的维数与变量个数相等，与延迟时间无关，这样做的好处就是避免了重构参数的选取问题，这方面的工作可见文献[102,108～111]，但当测量的变量个数小于原动力系统的实际维数时，这样做显然会失去某些信息，重构的空间不等价于原动力系统。文献[102]猜测最佳嵌入可能是某些变量和某些时间延迟的组合，为此后来在研究多变量时间序列的预测时，文献[112]提出了多变量时间序列延迟重构的思想，做以下时间延迟相空间重构

$$\boldsymbol{x}_n=(x_{1,n},x_{1,n-\tau_1},\cdots,x_{1,n-(m_1-1)\tau_1};x_{2,n},x_{2,n-\tau_2},\cdots,x_{2,n-(m_2-1)\tau_2};$$

$$\cdots;x_{L,n},x_{L,n-\tau_L},\cdots,x_{L,n-(m_L-1)\tau_L})$$

$$n=N_0,N_0+1,\cdots,N$$

$$N_0 = \max_{1\le l\le L}\{(m_l - 1)\tau_l + 1\} \tag{4.58}$$

式中，$\tau_l(l=1,2,\cdots,L)$和 $m_l(l=1,2,\cdots,L)$分别表示延迟时间间隔和嵌入维数，类似于 F. Takens 的延迟嵌入定理，只要 m_l 或 $m=\sum_{l=1}^{L} m_l$ 充分大，存在映射 $G: R^m \to R^m$，使

$$x_{n+1} = G(x_n) \tag{4.59}$$

也可以写成以下等价形式：

$$x_{l,n+1} = G_l(x_n),\ l = 1,2,\cdots,L \tag{4.60}$$

此时状态空间 $x_n \to x_{n+1}$的演化反映了原未知动力系统的演化，这意味着原动力系统吸引子的几何特征等价于重构的 m 维状态空间中吸引子的几何特征，因此原动力系统中任何微分或拓扑不变量可以在重构的状态空间中计算。

与单变量时间序列相空间重构一样，R^m 中 x_i 到 x_j 的距离记为 $\| x_i - x_j \|$，可以由 L_2 或 L_∞范数定义。

在按式(4.58)重构后，面临的问题是如何选取延迟时间间隔 $\tau_l(l=1,2,\cdots,L)$和嵌入维数 $m_l(l=1,2,\cdots,L)$，使式(4.59)或式(4.60)成立。

在多变量时间序列中，由于变量之间的性质不同，反应在数值上差异可能比较大，因此延迟重构参数 τ_l 的选取采用单变量时间序列中的方法，即对每一个分量时间序列$\{x_{l,n}\}_{n=1}^{N}$，分别确定 τ_l，具体可以用第二章中的自关联函数最小法或互信息最小法等方法确定，而嵌入维数 m_l 的确定比较复杂，下面具体讨论几种确定嵌入维数的方法。

4.3.1 利用预测误差最小法确定嵌入维数

预测误差最小法是 L. Y. Cao 等在文献[112]中提出的，后来文献[113]也利用了这种方法。其基本思想是利用式(4.59)中的映射 G 或式(4.60)中的 $G_l(l=1,2,\cdots,L)$的连续性。由于不同的 G_l 可能需要不同的最小嵌入维数(例如 Lorenz 系统中，$\frac{dx}{dt}$只依赖于 x 和 y 但不依赖于 z，而$\frac{dy}{dt}$依赖于 x、y 和 z)，因此实际计算中考虑式(4.60)中的 G_l。直觉上，由于 G_l 的连续性，当 m_l 为合适的嵌入维数时，若 x_n 与 x_j 充分靠近，则 $x_{l,n+1}$与 $x_{l,j+1}$也将充分靠近。

对已确定的时间延迟间隔 $\tau_l(l=1,2,\cdots,L)$及任意给定的嵌入维数 $m_l(l=1,2,\cdots,L)$，按式(4.58)重构向量 x_n，设 x_n 的最近邻点为 $x_{\eta(n)}$，即

$$\| x_n - x_{\eta(n)} \| = \min\{ \| x_n - x_j \| : j = N_0, N_0+1,\cdots,N, j \neq n\} \tag{4.61}$$

定义下面的一步预测误差

$$E(m_1,m_2,\cdots,m_L) = \frac{1}{N - N_0 + 1}\sum_{n=N_0}^{N} |x_{1,N+1} - x_{1,\eta(n)+1}| \tag{4.62}$$

对于较小的 $m_1,m_2,\cdots,m_L$，由于没有合适地嵌入，预测较差，所以 $E(m_1,m_2,\cdots,m_L)$较大；当 $m_1,m_2,\cdots,m_L$ 增加到合适的嵌入所需的最小值时，由函数 G_l 的连续性，$E(m_1,m_2,\cdots,m_L)$减少，如果系统有正的 Lyapunov 指数及噪声，则继续增加 $m_1,m_2,\cdots,m_L$ 会使 $E(m_1,m_2,\cdots,m_L)$也继续增加，因此可选取使 $E(m_1,m_2,\cdots,m_L)$最小的那组 $m_1,m_2,\cdots,m_L$ 为嵌入维数，即

$$(m_1^*,m_2^*,\cdots,m_L^*)=\arg\min\left\{E(m_1,m_2,\cdots,m_L)\,\middle|\,(m_1,m_2,\cdots,m_L)\in Z^L,\sum_{l=1}^{L}m_l\neq 0\right\} \tag{4.63}$$

式中，$Z^L=\prod_{l=1}^{L}Z$，Z 表示所有非负整数。

注：

1）代替一步预测误差式(4.62)，可以定义最大误差或均方根误差。

2）代替利用最近邻点(局部常数)定义一步预测，也可以定义多步预测或其他预测，然后再定义预测误差。但用最近邻点定义的一步预测计算快速简便，且不需要估计参数。

3）实际应用中，可限定在某一范围内搜索使 $E(m_1,m_2,\cdots,m_L)$达到最小的嵌入维数，即先确定某一最大维数 $D_{\max}$，考虑

$$(m_1^*,m_2^*,\cdots,m_L^*)=\arg\min\{E(m_1,m_2,\cdots,m_L)\mid 0\leqslant m_l\leqslant D_{\max},l=1,2,\cdots,L,\sum_{l=1}^{L}m_l\neq 0\} \tag{4.64}$$

4.3.2　利用虚假最近邻点法确定嵌入维数

在单变量时间序列相空间重构中，虚假最近邻点法是一种比较理想的确定嵌入维数的算法，我们在文献[114]中把这种算法推广到了多变量的情况。

设 $\boldsymbol{x}_n$ 的最近邻点为 $\boldsymbol{x}_{\eta(n)}$，即

$$\|\boldsymbol{x}_n-\boldsymbol{x}_{\eta(n)}\|^{(m_1,\cdots,m_l,\cdots,m_L)}=\min\{\|\boldsymbol{x}_n-\boldsymbol{x}_j\|:j=N_0,N_0+1,\cdots,N,j\neq n\} \tag{4.65}$$

当维数从 m 增加到 $m+1$ 时，如 m_l 变为 m_l+1，则它们之间的距离变为 $\|\boldsymbol{x}_n-\boldsymbol{x}_{\eta(n)}\|^{(m_1,\cdots,m_l+1,\cdots,m_L)}$，若 $\|\boldsymbol{x}_n-\boldsymbol{x}_{\eta(n)}\|^{(m_1,\cdots,m_l+1,\cdots,m_L)}$比 $\|\boldsymbol{x}_n-\boldsymbol{x}_{\eta(n)}\|^{(m_1,\cdots,m_l,\cdots,m_L)}$大很多，可以认为是由于高维吸引子中两个不相邻的点在投影到低维轨线上时变成相邻的两点造成的，因此这样的邻点是虚假的。具体地说，当

$$\frac{\|\boldsymbol{x}_n-\boldsymbol{x}_{\eta(n)}\|^{(m_1,\cdots,m_l+1,\cdots,m_L)}-\|\boldsymbol{x}_n-\boldsymbol{x}_{\eta(n)}\|^{(m_1,\cdots,m_l,\cdots,m_L)}}{\|\boldsymbol{x}_n-\boldsymbol{x}_{\eta(n)}\|^{(m_1,\cdots,m_l,\cdots,m_L)}}\geqslant R_T \tag{4.66}$$

时，$\boldsymbol{x}_{\eta(n)}$为 $\boldsymbol{x}_n$ 的虚假最近邻点。

对实测时间序列，$(m_1,\cdots,m_l,\cdots,m_L)$从$(1,\cdots,1,\cdots,1)$开始，每次 m_l 增加 1，

直到虚假最近邻点的比例小于某个 δ 或虚假最近邻点不再随着 m 的增加而减少时，可以认为吸引子几何结构完全打开，此时的 $(m_1,\cdots,m_l,\cdots,m_L)$ 为嵌入维数。根据经验通常取定义 $R_T\in[10,50]$，$\delta=5\%$。

上述算法中的两个关键阈值参数 R_T 和 δ 对整个算法的性能和有效性十分重要。R_T 用来判别邻点是否为虚假，过大的选择会造成无法发现虚假邻点，而过小的选择却又会受到时间序列中噪声的干扰，因此选择合适的 R_T 是一个较困难的问题。δ 是用来标记虚假邻点率的，理论上 $\delta=0$ 表示所有虚假邻点的消失，但是实际上由于时间序列在观测过程中受到噪声的干扰 δ 无法收敛到 0，因此多凭借经验根据观测时间序列质量的好坏预定一个阈值。

4.3.3　虚假最近邻点法确定嵌入维数算法的改进

针对虚假最近邻点算法的不足，我们在文献[115]中给出了一个更加实用的改进算法，避免了阈值参数选取的主观性。

对 L 变量时间序列 $\{x_{l,n}\}_{n=1}^{N}$，$l=1,2,\cdots,L$，作式(4.58)的相空间重构，设 $\boldsymbol{x}_n$ 的最近邻点为 $\boldsymbol{x}_{\eta(n)}$，即

$$\|\boldsymbol{x}_n-\boldsymbol{x}_{\eta(n)}\|=\min\{\|\boldsymbol{x}_n-\boldsymbol{x}_j\|:j=N_0,N_0+1,\cdots,N,j\neq n\}\quad(4.67)$$

定义

$$e(n;m_1,\cdots,m_l,\cdots,m_L)=\frac{\frac{1}{L}\sum_{l=1}^{L}\|\boldsymbol{x}_n(m_1,\cdots,m_l+1,\cdots,m_L)-\boldsymbol{x}_{\eta(n)}(m_1,\cdots,m_l+1,\cdots,m_L)\|}{\|\boldsymbol{x}_n(m_1,\cdots,m_l,\cdots,m_L)-\boldsymbol{x}_{\eta(n)}(m_1,\cdots,m_l,\cdots,m_L)\|}\quad(4.68)$$

这里 $n=N_0,N_0+1,\cdots,N$，进一步定义

$$E(m_1,\cdots,m_l,\cdots,m_L)=\frac{1}{N-N_0+1}\sum_{n=N_0}^{N}e(n;m_1,\cdots,m_l,\cdots,m_L)\quad(4.69)$$

记集合

$$U(m)=\{E(m_1,\cdots,m_l,\cdots,m_L)\mid m_1+\cdots+m_l+\cdots+m_L=m\}\quad(4.70)$$

且定义

$$E(m)=\frac{1}{|U(m)|}\sum_{E(m_1,\cdots,m_l,\cdots,m_L)\in U(m)}E(m_1,\cdots,m_l,\cdots,m_L)\quad(4.71)$$

其中 $|U(m)|$ 是集合 $U(m)$ 所含元素的总数。为了观察收敛情况，定义

$$E_1(m)=\frac{E(m+1)}{E(m)}\quad(4.72)$$

如果 $m>m_0$ 时，$E_1(m)$ 平稳，则 m_0+1 为整个重构的相空间的嵌入维数。进一步计算使得 $E(m_1,\cdots,m_l,\cdots,m_L)$ 最小化的 $m_1^*,\cdots,m_l^*,\cdots,m_L^*$，即

$E(m_1^*,\cdots,m_l^*,\cdots,m_L^*)$

$$= \min\{E(m_1,\cdots,m_l,\cdots,m_L) \mid E(m_1,\cdots,m_l,\cdots,m_L) \in U(m_0+1)\} \quad (4.73)$$

则 $m_1^*,\cdots,m_l^*,\cdots,m_L^*$ 是由 L 变量时间序列 $\{x_{l,n}\}_{n=1}^N, l=1,2,\cdots,L$ 重构相空间后各子空间的嵌入维数。

该算法可以在计算全局相空间重构维数的过程中，同时计算各独立系统的局部相空间重构维数。由于简单的虚假邻点算法在计算嵌入维数时，使用了固定的距离判别参数 R_T，无法适应数据异常波动的情况，特别是当时间序列中含有噪声信号时 R_T 的选择非常困难，针对这个缺陷，改进算法通过使用式(4.68)，引入函数 $e(n;m_1,\cdots,m_l,\cdots,m_L)$ 代替了原来的常数 R_T，这里 $e(n;m_1,\cdots,m_l,\cdots,m_L)$ 是一个依赖于 n 和 m 的函数，该函数可以使判别距离随着嵌入空间的维数和使用的参考向量不同而变化，从而提高了适应异常数据波动的能力。另外，虚假邻点算法的收敛判别是通过使用固定参数 δ 确定的，显然在不同的应用场合提前给定一个收敛阈值是不合适的，如果实际数据中含有高级别的噪声信号，则收敛阈值需要放大到25%，而当信号比较干净的时候，这个阈值又需要减少到5%；为了解决这个问题，改进算法通过式(4.72)，给出了一个多变量环境下搜索嵌入维数时的收敛变化曲线，通过观测 $E_1(m)$ 随嵌入维数增加的变化曲线，我们可以找出 $E_1(m)$ 不再增加或者变化的嵌入维数 m_0，则 m_0+1 就是所求的最佳嵌入维数。该算法经过测试对离散型或者连续型非线性动力系统产生的非线性时间序列均可以有效地寻找出最佳重构嵌入维数参数。在有噪声干扰的情况下，仍然可以给出比较接近真实情况的估计值。作为计算的一个最终结果，式(4.73)还可以给出各个参与重构变量所需要使用的局部相空间重构维数。

由此可以看出，在多变量环境下该改进算法避免使用这两个关键参数 R_T 和 δ，使得计算工作更加简单，改进算法在搜索全局最佳嵌入维数的同时还可以给出最佳的局部嵌入空间维数组合，简化了多变量环境下相空间的重构工作。

4.3.4 嵌入维数算法的仿真计算

例 4.4 考虑以下 Lorenz 系统

$$\begin{cases} \dfrac{dx}{dt} = \sigma(y-x) \\ \dfrac{dy}{dt} = x(r-z)-y \\ \dfrac{dz}{dt} = xy-bz \end{cases} \quad (4.74)$$

其中 $\sigma=10, r=28, b=\dfrac{8}{3}$。取初值 $x_0=15.34, y_0=13.68, z_0=37.91$，利用四阶 Runge Kute 法积分分别计算 x,y,z，积分步长为 $h=0.04$，计算过程中精度为 0.0001，得长度 $N=2000$ 的三个变量的时间序列 $\{x_n\}_{n=1}^N$，$\{y_n\}_{n=1}^N$，$\{z_n\}_{n=1}^N$，对前两

个变量构成的双变量时间序列$\{x_n\}_{n=1}^{N}$和$\{y_n\}_{n=1}^{N}$，利用互信息最小可得它们的最佳延迟时间间隔为$\tau_1=\tau_2=4$，按式(4.58)做以下双变量时间延迟重构

$$x_n=(x_n,x_{n-4},\cdots,x_{n-4(m_1-1)};y_n,y_{n-4},\cdots,y_{n-4(m_2-1)})$$

$$n=N_0,N_0+1,\cdots,N;\ N_0=4[\max(m_1,m_2)-1]+1 \qquad (4.75)$$

若记当m_1变成m_1+1时按式(4.66)计算的虚假最近邻点数为K_1，当m_2变成m_2+1时按式(4.66)计算的虚假最近邻点数为K_2，则虚假最近邻点数随阈值及嵌入维数的变化情况可见表4.1。

从表4.1可见，$m_1=1,m_2=2$或$m_1=2,m_2=1$时，虚假最近邻点数为K_1和K_2都很小，几乎接近于零，在总点数中的比例已很小，因此可以认为它们是最小嵌入维数。

表4.1 虚假最近邻点数随阈值及嵌入维数的变化情况

阈值 R_T	嵌入维数 m_1	嵌入维数 m_2	虚假最近邻点数 K_1	虚假最近邻点数 K_2
20	1	1	17	12
	1	2	0	1
	2	1	3	0
25	1	1	13	10
	1	2	0	1
	2	1	3	0
30	1	1	13	7
	1	2	0	0
	2	1	2	0

例4.5 考虑如下两个不完全相同的Rossler方程的耦合系统

$$\begin{cases}\dfrac{dx_1}{dt}=-\omega_1y_1-z_1+\varepsilon(x_2-x_1)\\ \dfrac{dy_1}{dt}=\omega_1x_1+0.15y_1\\ \dfrac{dz_1}{dt}=0.2+z_1(x_1-10)\\ \dfrac{dx_2}{dt}=-\omega_2y_2-z_2+\varepsilon(x_1-x_2)\\ \dfrac{dy_2}{dt}=\omega_2x_2+0.15y_2\\ \dfrac{dz_2}{dt}=0.2+z_2(x_2-10)\end{cases} \qquad (4.76)$$

其中$\omega_1=0.99$，$\omega_2=0.95$。文献[113，116]的计算过程，系统初始状态为$x_{1,0}=0.1,y_{1,0}=0.2,z_{1,0}=0.3,x_{2,0}=y_{2,0}=0,z_{2,0}=15$，这里取$\varepsilon=0.05$和0.50，耦合参数$\varepsilon$对应耦合系统的间歇延迟同步和完全同步两种状态。

采用四阶 Runge-Kuta 积分法分别计算变量 $x_1, y_1, z_1, x_2, y_2, z_2$ 的 15 000 个观测值，并取后 2000 个数据作为试验对象。分别对变量 x_1 作单变量时间序列重构，对变量 x_1, x_2 作两变量时间序列重构，对变量 x_1, x_2, y_1 作三变量时间序列重构，对变量 x_1, x_2, y_1, y_2 作四变量时间序列重构。延迟时间间隔采用互信息法计算，对变量 x_1 作单变量时间序列重构时，嵌入维数采用 2.2.1 节中改进的虚假邻点法计算；对变量 x_1, x_2 作两变量时间序列重构时，对变量 x_1, x_2, y_1 作三变量时间序列重构时，对变量 x_1, x_2, y_1, y_2 作四变量时间序列重构时，嵌入维数采用 4.3.3 节中改进的虚假邻点法计算，计算结果见表 4.2。

表 4.2　不同耦合强度下各种重构方式的重构参数

重构参数 m/τ ＼ 耦合强度 重构变量	0.05	0.50
单变量 x_1 重构	6/17	7/14
两变量 x_1, x_2 重构	4/17,3/14	6/14,1/14
三变量 x_1, x_2, y_1 重构	1/17,4/14,2/11	6/14,1/14,1/14
四变量 x_1, x_2, y_1, y_2 重构	1/17,4/14,2/11,1/16	1/14,4/14,1/14,1/14

例 4.6　考虑如下 Rossler 方程和 Hyper Rossler 方程的耦合系统

$$\left.\begin{aligned}
\frac{dx_1}{dt} &= -\omega y_1 - z_1 + \varepsilon(x_2 - x_1) \\
\frac{dy_1}{dt} &= \omega x_1 + 0.15 y_1 \\
\frac{dz_1}{dt} &= 0.2 + z_1(x_1 - 10) \\
\frac{dx_2}{dt} &= w_2 + 0.25 x_2 + z_2 + \varepsilon(x_1 - x_2) \\
\frac{dy_2}{dt} &= 3 + y_2 w_2 \\
\frac{dz_2}{dt} &= -0.5 y_2 + 0.05 z_2 \\
\frac{dw_2}{dt} &= -x_2 - y_2
\end{aligned}\right\} \tag{4.77}$$

其中 $\omega = 0.925$。系统初始状态为 $x_{1,0}=0.1, y_{1,0}=0.2, z_{1,0}=0.3, x_{2,0}=y_{2,0}=0$, $z_{2,0}=15, w_{2,0}=-20$，这里取 $\varepsilon=0.008$ 和 0.012，耦合参数 ε 对应耦合系统的非同步和同步演化。

采用四阶 Runge-Kuta 积分法分别计算变量 $x_1, y_1, z_1, x_2, y_2, z_2, w_2$ 的 15 000

个观测值，并取后 2000 个数据作为试验对象。分别对变量 x_1 作单变量时间序列重构，对变量 x_1, x_2 作两变量时间序列重构，对变量 x_1, x_2, w_2 作三变量时间序列重构，对变量 x_1, x_2, w_2, y_1 作四变量时间序列重构。延迟时间间隔采用互信息法计算，对变量 x_1 作单变量时间序列重构时，嵌入维数采用 2.2.1 节中改进的虚假邻点法计算；对变量 x_1, x_2 作两变量时间序列重构时，对变量 x_1, x_2, w_2 作三变量时间序列重构时，对变量 x_1, x_2, w_2, y_1 作四变量时间序列重构时，嵌入维数采用 4.3.3 节中改进的虚假邻点法计算，计算结果见表 4.3。

表 4.3　不同耦合强度下各种重构方式的重构参数

重构参数 m/τ ＼ 耦合强度 ／ 重构变量	0.008	0.012
单变量 x_1 重构	7/15	8/15
两变量 x_1, x_2 重构	5/15,3/12	2/15,6/14
三变量 x_1, x_2, w_2 重构	2/15,1/12,3/15	3/15,2/14,1/18
四变量 x_1, x_2, w_2, y_1 重构	3/15,5/12,1/15,1/15	3/15,2/14,5/18,1/13

上述多变量相空间重构参数的选择是否合适，将在 4.4 节中通过计算几何不变量来说明。

4.4　多变量时间序列重构相空间中几何不变量的计算

在多变量时间序列重构的相空间中，几何不变量的计算与单变量时间序列重构的相空间中类似，这里我们将把 P. Grassberger 和 I. Procaccia 给出的 G-P 关联维数推广到广义关联维数，并讨论在时间序列长度较短的情况下，最大 Lyapunov 指数的计算方法。由于多变量时间序列包含更多的信息，因此可以预期在计算几何不变量时对时间序列的长度要求会比单变量时间序列的情形有所降低。

4.4.1　广义关联维数的计算

在按式(4.58)重构的 m 维空间中，$\sum_{j=N_0, j\neq i}^{N} H(r-\| \boldsymbol{x}_i-\boldsymbol{x}_j \|)$ 表示除 $\boldsymbol{x}_i$ 本身外到 $\boldsymbol{x}_i$ 的距离小于 r 的 $\boldsymbol{x}_j$ 的点数，其中 $H(\cdot)$ 是 Heavside 函数，定义按上述方式重构的有 L 个变量的时间序列 $\{x_{l,n}\}_{n=1}^{N}$ $(l-1,2,\cdots,L)$ 的 q 阶广义关联积分为

$$C_q^L(r)=C_q^L(r;\tau_1,\cdots,\tau_L;m_1,\cdots,m_L)$$

$$= \left\{ \frac{1}{N-N_0+1} \sum_{i=N_0}^{N} \left[\frac{1}{N-N_0} \sum_{j=N_0, j\neq i}^{N} H(r - \| x_i - x_j \|) \right]^{q-1} \right\}^{\frac{1}{q-1}} \quad (4.78)$$

其中 q 是正整数，$q\neq 1$，特别是 $q=2$ 时就是通常的关联积分。当 $q=1$ 时，即一阶广义关联积分定义为

$$\ln C_1^L(r) = \lim_{q\to 1} \ln C_q^L(r)$$

$$= \lim_{q\to 1} \frac{1}{q-1} \ln \frac{1}{N-N_0+1} \sum_{i=N_0}^{N} \left[\frac{1}{N-N_0} \sum_{j=N_0, j\neq i}^{N} H(r - \| x_i - x_j \|) \right]^{q-1}$$

$$= \frac{1}{N-N_0+1} \sum_{i=N_0}^{N} \ln \frac{1}{N-N_0} \sum_{j=N_0, j\neq i}^{N} H(r - \| x - x_j \|) \quad (4.79)$$

其中 $C_q^L(r)$ 描述了距离小于 r 的对点数的分布情况。在数据无噪声、无限精确且无限长这种理想状况下，当 $q=2$，$N\to\infty$ 时，$C_q^L(r)\to C(r)=p\{\| x_i - x_j \| < r\}$，即 m 维空间中两个随机选取的点 x_i 与 x_j 在同一半径为 r 的球内的概率，此时关联维数的严格定义为[21,22]

$$d_2 = \lim_{m\to\infty} \lim_{r\to 0} \frac{\ln C(r)}{\ln r} \quad (4.80)$$

同理可以定义广义关联维数为

$$d_q = \frac{1}{q-1} \lim_{m\to\infty} \lim_{r\to 0} \lim_{N\to\infty} \frac{\ln C_q^L(r)}{\ln r} \quad (4.81)$$

它们是刻画系统复杂程度的一个几何不变量。但在实际应用中，由于数据有噪声及测量误差，且不可能无限长，因此只能找关联维数的近似值来刻画系统的复杂性。在实测有限长时间序列中，若 $r > \max\limits_{N_0\leqslant i<j\leqslant N} \| x_i - x_j \|$，一切点对的距离都不会超过它，因而 $C_q^L(r)\equiv 1$；若 $r < \min\limits_{N_0\leqslant i<j\leqslant N} \| x_i - x_j \|$，则 $C_q^L(r)\equiv 0$。因此 r 既不能选得太大，也不能选得太小，太大或太小的 r 都不能反映系统内部的性质，同样由于受到数据长度的限制，m 也不能取得太大。如果在 r 的某一区间段内，有

$$C_q^L(r) \propto r^{(q-1)d_q} \quad (4.82)$$

则称 d_q 是 q 阶广义关联维数，$q=2$ 时就是通常的 G-P 关联维数 d。可以认为这样定义的 d_q 就是近似刻画多变量时间序列描述的复杂系统复杂程度的某种维数。

下面的问题是式(4.82)什么时候成立？一种思想是让嵌入维数 m 变化，当 $m<d_q$ 时，由于重构相空间的维数不够高，以致不能分辨吸引子的结构，因此，此时双对数 $\ln C_q^L(r)$ $\ln r$ 图的斜率近似取嵌入维数；当 m 增加时，重构相空间中吸引子的分辨能力改善，m 增加到斜率达到一个平稳时期为止，此时双对数 $\ln C_q^L(r)$ $\ln r$ 图中可找到平行直线区域，这些直线的斜率与 m 的依赖性消失，该斜率即为所求的广义关联维数 d_q。对单变量无噪声无限长时间序列，文献[11]证明了只要 $m=\mathrm{ceil}(d_2)$[其中 $\mathrm{ceil}(d_2)$ 是大于等于 d_2 的最小整数]就出现平稳时期，但对有限长具

噪声的时间序列，平稳时期开始的 m 要大于 $\text{ceil}(d_2)$。另一种思想就是假定按前面的方法重构了相空间，并确定了嵌入维数，此时式(4.82)应在 r 的某区间内成立，则可求出 d_q。

在实际应用中，确定了平稳时期后，直线的斜率可由不同方法计算。例如，当 r_1,r_2 在满足式(4.82)的 r 的某区间段内，则

$$d_q=\frac{\ln C_q^L(r_2)-\ln C_q^L(r_1)}{\ln r_2-\ln r_1} \tag{4.83}$$

或在满足式(4.82)的 r 的某区间段内取 K 个点 $r_1,r_2,\cdots,r_K$，由最小二乘法，得

$$d_q=\frac{K\sum_{k=1}^{K}\ln r_k\ln C_q^L(r_k)-\sum_{k=1}^{K}\ln r_k\sum_{k=1}^{K}\ln C_q^L(r_k)}{K\sum_{k=1}^{K}(\ln r_k)^2-\left(\sum_{k=1}^{K}\ln r_k\right)} \tag{4.84}$$

实测时间序列在找双对数 $\ln C_q^L(r)$-$\ln r$ 图的线性部分时，由于连续采样点之间的相关性，会造成"肩峰"效应，为了消除这种"肩峰"效应，参考单变量时间序列关联积分的处理方式，把广义关联积分式(4.77)和式(4.78)修改为

$$C_q^L(r,\omega)=$$

$$\left\{\frac{1}{N-N_0-\omega}\sum_{i=N_0+\omega+1}^{N}\left[\frac{1}{N-N_0-2\omega+2}\left(\sum_{j=N_0}^{i-\omega}+\sum_{j=i+\omega}^{N}\right)H(r-\|x_i-x_j\|)\right]^{q-1}\right\}^{\frac{1}{q-1}},\quad q\neq 1 \tag{4.85}$$

$$\ln C_1^L(r,\omega)=\frac{1}{N-N_0-\omega}\sum_{i=N_0+\omega+1}^{N}\ln\frac{1}{N-N_0-2\omega+2}\left(\sum_{j=N_0}^{i-\omega}+\sum_{j=i+\omega}^{N}\right)H(r-\|x_i-x_j\|) \tag{4.86}$$

说明：

1) 在数据无噪声、无限精确及无限长的理想状况下，q 阶广义关联维数与 R^m 中所采用的范数无关，这是由于

$$\|x_i-x_j\|_\infty\leqslant\|x_i-x_j\|_2\leqslant\sqrt{m}\,\|x_i-x_j\|_\infty \tag{4.87}$$

因此

$$H(r-\|x_i-x_j\|_\infty)\geqslant H(r-\|x_i-x_j\|_2)\geqslant H(r-\sqrt{m}\,\|x_i-x_j\|_\infty)$$

$$=H\left(\frac{r}{\sqrt{m}}-\|x_i-x_j\|_\infty\right) \tag{4.88}$$

则

$$C_{q,\mathrm{l}\cdot\mathrm{l}_\infty}^L(r)\geqslant C_{q,\mathrm{l}\cdot\mathrm{l}_2}^L(r)\geqslant C_{q,\mathrm{l}\cdot\mathrm{l}_\infty}^L\left(\frac{r}{\sqrt{m}}\right) \tag{4.89}$$

从而由式(4.81)及极限计算的夹逼定理知

$$d_{q,\mathrm{l}\cdot\mathrm{l}_\infty}=d_{q,\mathrm{l}\cdot\mathrm{l}_2} \tag{4.90}$$

2）不同研究对象的观测时间序列数据的量级差别往往很大，在确定计算广义关联维数的尺度区域时，r 的选择差异较大，因此造成的累积误差可能也很大，为此可对数据作标准化处理。设观测获得的多变量时间序列为 $\{x_{l,n}\}_{n=1}^{N}(l=1,2,\cdots,L)$，定义

$$x_{l,n}^{*}=\frac{x_{l,n}-x_{\min}}{x_{\max}-x_{\min}},\ n=1,2,\cdots,N;\ l=1,2,\cdots,L \tag{4.91}$$

其中 $x_{\min}=\min\{x_{l,n}\mid l=1,2,\cdots,L;\ n=1,2,\cdots,N\}$，$x_{\max}=\max\{x_{l,n}\mid l=1,2,\cdots,L;\ n=1,2,\cdots,N\}$，则 $0\leqslant x_{l,n}^{*}\leqslant 1$，无论是 L^2 范数还是 L^{∞} 范数作为距离定义，都有

$$\| x_i^{*}-x_j^{*}\| =\frac{1}{x_{\max}-x_{\min}}\| x_i-x_j\| \tag{4.92}$$

因此

$$C_q^{*,L}(r)=C_q^L[(x_{\max}-x_{\min})r] \tag{4.93}$$

则

$$C_q^{*,L}(r)\propto C_q^L(r)\propto r^{(q-1)d_q} \tag{4.94}$$

即 $\{x_{l,n}\}_{n=1}^{N}(l=1,2,\cdots,L)$ 与 $\{x_{l,n}^{*}\}_{n=1}^{N}(l=1,2,\cdots,L)$ 有相同的广义关联维数，而对 $\{x_{l,n}^{*}\}_{n=1}^{N}(l=1,2,\cdots,L)$ 来说，尺度区域内的 r 总满足

$$0\leqslant\frac{\min\limits_{N_0\leqslant i<j\leqslant N}\| x_i-x_j\|}{x_{\max}-x_{\min}}\leqslant r\leqslant\frac{\max\limits_{N_0\leqslant i<j\leqslant N}\| x_i-x_j\|}{x_{\max}-x_{\min}} \tag{4.95}$$

若取 L^{∞} 范数，则 r 总满足

$$0\leqslant\frac{\min\limits_{N_0\leqslant i<j\leqslant N}\| x_i-x_j\|}{x_{\max}-x_{\min}}\leqslant r\leqslant 1 \tag{4.96}$$

在双对数 $\ln C_q^L(r)$-$\ln r$ 图上，$\ln r$ 可在 $[-A,0]$ 之间选取，A 是某一正数。

下面通过一个具体的例子来说明多变量时间序列相空间重构中计算广义关联维数的精确程度。对例 4.4 的 Lorenz 系统的前两个变量构成的双变量时间序列 $\{x_n\}_{n=1}^{N}$ 和 $\{y_n\}_{n=1}^{N}$，按例 4.4 中的计算，取 $m_1=1$，$m_2=2$ 作为最小嵌入维数，$\tau_1=\tau_2=4$ 作为最佳延迟时间间隔，按式(4.58)做以下双变量时间延迟重构

$$\boldsymbol{x}_n=(x_n,y_n,y_{n-4}),\ n=5,6,\cdots,N \tag{4.97}$$

取时间序列长度 $N=16\ 384$，r 在 $\min\limits_{5\leqslant i<j\leqslant N}\| \boldsymbol{x}_i-\boldsymbol{x}_j\| =\exp(-5.546\ 778\ 73)$ 和 $\max\limits_{5\leqslant i<j\leqslant N}\| \boldsymbol{x}_i-\boldsymbol{x}_j\| =\exp(3.952\ 235\ 53)$ 之间取值，取 201 个均匀的点得到了图 4.8 所示的 $\ln C_q^L(r)$-$\ln r$ 图。

从图 4.8 中可看出，当 r 较小时，$C_1^2(r)$ 产生了“肩峰”效应，但当 r 在 $\exp(-1.034\ 746\ 95)$ 和 $\exp(0.770\ 065\ 76)$ 之间取值时，达到了平稳时期，即可找到双对数 $\ln C_q^L(r)$-$\ln r$ 图的尺度区域(scaling region)，在此区域内，图形几乎成直线，因此，由式(4.84)，取 $K=39$ 个均匀的点，经计算得 $d_1=2.009\ 86$，$d_2=$

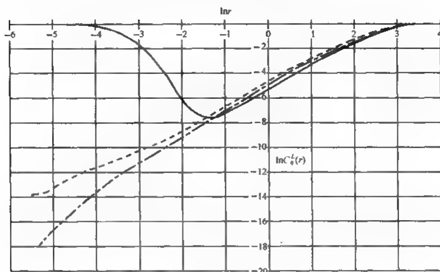


图 4.8　Lorenz 系统的 $\ln C_q^L(r)$-$\ln r$ 图

实线表示 $q=1$；点画线表示 $q=2$；虚线表示 $q=3$

2.070 68，$d_3=2.003\ 25$，这与文献[26]中的 $d_2\approx2.04$ 以及文献[117]中的 $d_2\approx$ 2.01 非常接近，但这里所用的时间序列长度要短得多。

4.4.2　小数据量情况下最大 Lyapunov 指数的计算

最大 Lyapunov 指数 λ_1 定量地描述系统相空间相邻轨道呈指数发散或收敛的性质，是描述混沌动力学特性的重要参数之一，$\lambda_1>0$ 是系统具有混沌特性的重要指标。当系统的模型已知时，可以利用相空间的切向量精确计算最大 Lyapunov 指数。当复杂系统只能获得观测时间序列时，为了计算最大 Lyapunov 指数，A. Wolf 提出了轨道跟踪法，随后 M. T. Rosenstein 等进行了改进，但仍存在精度不高、受噪声影响大、计算量大等问题[118]。为此，文献[119]在深入研究相空间重构技术和轨道跟踪技术的基础上提出了一种比较稳健的计算方法。这些方法都是对观测的单变量时间序列进行处理的，但在实际问题中，被观测的复杂系统往往有多个变量描述，通过观测或实验手段可以获得多变量时间序列，理论上多变量时间序列比单变量时间序列包含更多的关于原复杂系统的信息，采用多变量时间序列计算得到的最大 Lyapunov 指数应更加真实。同时，实际问题中通过观测或实验获得的时间序列可能不是很长，较短的多变量时间序列是否也能获得较精确的最大 Lyapunov 指数，这是一个很有意义的问题。

下面将把文献[118, 119]中对小数据量单变量时间序列的最大 Lyapunov 指

数算法推广到多变量时间序列的情况下。由于小数据量方法与 A. Wolf 方法一样是基于重构相空间进行计算的，因此首先使用多变量时间序列相空间重构参数的选择算法寻找重构参数并利用计算结果进行重构，进一步利用重构相空间采用小数据量方法计算 Lyapunov 指数，并通过 6 维和 7 维耦合 Rossler 混沌系统产生的多变量数据仿真计算检验了算法的有效性。

设 $\{x_{l,n}\}_{n=1}^{N}(l=1,2,\cdots,L)$ 为观测的 L 变量的时间序列，做式(4.58)的相空间重构。对每个点 x_j 寻找其最近邻点 $x_{\eta(j)}$，这两个点之间必须有短暂的分离，以保证两个点沿着不同的轨道运行。定义分离间隔

$$\omega=\frac{\max\tau_l}{\Delta t},\ l=1,2,\cdots,L \tag{4.98}$$

假定 $d_j(0)$ 为 x_j 到其最近邻点 $x_{\eta(j)}$ 的距离，即

$$d_j(0)=\|x_j-x_{\eta(j)}\|,\ |j-\eta(j)|>\omega \tag{4.99}$$

对相空间中的每个点 x_j，计算出其最近邻点在 i 步前向演化后的距离

$$d_j(i)=\|x_{j+i}-x_{\eta(j)+i}\|,\ i=N_0,N_0+1,\cdots,N \tag{4.100}$$

假定 x_j 的最近邻点以最大 Lyapunov 指数的速率发散，即

$$d_j(i)=d_j(0)\times e^{\lambda_1(i\times\Delta t)} \tag{4.101}$$

两边取对数，得

$$\ln d_j(i)=\ln d_j(0)+\lambda_1(i\times\Delta t) \tag{4.102}$$

在 $\ln d_j(i)$-$i\times\Delta t$ 图中，利用最小二乘法拟合得到最大 Lyapunov 指数 λ_1 为

$$\lambda_1=\frac{1}{\sum i^2}[i\times y(i)] \tag{4.103}$$

式中，$y(i)=\frac{1}{p\Delta t}\sum_{j=1}^{p}\ln d_j(i)$；$p$ 为非零的 $d_j(i)$ 的个数。

为了说明在时间序列长度较短的情况下，上述计算最大 Lyapunov 指数方法的有效性，同时也为了说明多变量时间序列相空间重构的必要性，下面通过两个耦合混沌系统进行仿真分析。

对例 4.5 中的两个不完全相同的 Rossler 方程的耦合系统，在 $\varepsilon=0.05$ 和 0.50 时，最大 Lyapunov 指数 λ_1 的理论值分别为 0.076 804 和 0.095 137。按例 4.5 的计算所得的重构参数进行相空间重构，并按上述方法计算其最大 Lyapunov 指数，结果见表 4.4 和表 4.5。

表 4.4 和表 4.5 中的第二列表示例 4.5 中计算的相空间重构参数，第三列表示按上面的方式计算的最大 Lyapunov 指数 λ_1，第四列表示计算结果与理论值的偏差，用百分比表示。

表 4.4　耦合系统(4.76)在 $\varepsilon=0.05$ 时的最大 Lyapunov 指数

重构变量	重构参数 m/τ	λ_1 的计算值	偏差
单变量 x_1 重构	6/17	0.061 173	−20.35%
两变量 x_1,x_2 重构	4/17,3/14	0.066 882	−12.92%
三变量 x_1,x_2,y_1 重构	1/17,4/14,2/11	0.070 499	−8.21%
四变量 x_1,x_2,y_1,y_2 重构	1/17,4/14,2/11,1/16	0.071 536	−6.86%

表 4.5　耦合系统(4.76)在 $\varepsilon=0.50$ 时的最大 Lyapunov 指数

重构变量	重构参数 m/τ	λ_1 的计算值	偏差
单变量 x_1 重构	7/14	0.073 55	−22.69%
两变量 x_1,x_2 重构	6/14,1/14	0.083 21	12.54%
三变量 x_1,x_2,y_1 重构	6/14,1/14,1/14	0.0858	−9.81%
四变量 x_1,x_2,y_1,y_2 重构	1/14,4/14,1/14,1/14	0.090 84	−4.51%

同样对例 4.6 中 Rossler 方程和 Hyper Rossler 方程的耦合系统，在 $\varepsilon=0.008$ 和 0.012 时，最大 Lyapunov 指数 λ_1 的理论值分别为 0.104 558 和 0.128 993。按例 4.6 的计算所得的重构参数进行相空间重构，并按上述方法计算其最大 Lyapunov 指数，结果见表 4.6 和表 4.7。

表 4.6　耦合系统(4.77)在 $\varepsilon=0.008$ 时的最大 Lyapunov 指数

重构变量	重构参数 m/τ	λ_1 的计算值	偏差
单变量 x_1 重构	7/15	0.042 245	−59.60%
两变量 x_1,x_2 重构	5/15,3/12	0.046 174	−55.84%
三变量 x_1,x_2,w_1 重构	2/15,1/12,3/15	0.084 368	−19.31%
四变量 x_1,x_2,w_2,y_1 重构	3/15,5/12,1/15,3/1	0.103 549	−0.97%

表 4.7　耦合系统(4.77)在 $\varepsilon=0.012$ 时的最大 Lyapunov 指数

重构变量	重构参数 m/τ	λ_1 的计算值	偏差
单变量 x_1 重构	8/15	0.083 325	−35.40%
两变量 x_1,x_2 重构	2/15,6/14	0.098 919	−23.31%
三变量 x_1,x_2,w_2 重构	3/15,2/14,1/18	0.106 871	−17.15%
四变量 x_1,x_2,w_2,y_1 重构	3/15,2/14,5/18,1/13	0.121 149	−6.08%

表 4.4～表 4.7 的计算结果表明，采用多变量时间序列计算最大 Lyapunov 指数明显优于采用单变量时间序列时的计算结果。相对于理论计算值的偏差，在采用 4 个变量时偏差可以下降到 −7%以内。该仿真计算表明对较短的时间序列在多变量情况下计算最大 Lyapunov 指数是可行的。

为了进一步验证和说明时间序列长度对计算最大 Lyapunov 指数的影响，对例 4.5 和例 4.6 中的两个系统，按前面的相空间重构参数，在单变量重构时，取 2000 个数据，在多变量重构时，时间序列的长度取 $2000/l$，其中 l 是变量的个数，即随着变量数量的增加而减少时间序列的长度，计算结果分别见表 4.8～表 4.11。

表 4.8 耦合系统(4.76)在 $\varepsilon=0.05$ 时的最大 Lyapunov 指数的比较

变量个数 l	时间序列长度为 2000 时计算的 λ_1	时间序列长度为 $2000/l$ 计算的 λ_1	λ_1 的理论值
1	0.061 173	0.061 173	0.076 804
2	0.066 882	0.080 560	0.076 804
3	0.070 499	0.084 590	0.076 804
4	0.071 536	0.082 160	0.076 804

表 4.9 耦合系统(4.76)在 $\varepsilon=0.50$ 时的最大 Lyapunov 指数的比较

变量个数 l	时间序列长度为 2000 时计算的 λ_1	时间序列长度为 $2000/l$ 计算的 λ_1	λ_1 的理论值
1	0.073 547	0.073 547	0.095 137
2	0.083 211	0.082 670	0.095 137
3	0.085 802	0.085 450	0.095 137
4	0.090 843	0.083 840	0.095 137

表 4.10 耦合系统(4.77)在 $\varepsilon=0.008$ 时的最大 Lyapunov 指数的比较

变量个数 l	时间序列长度为 2000 时计算的 λ_1	时间序列长度为 $2000/l$ 计算的 λ_1	λ_1 的理论值
1	0.042 245	0.042 245	0.104 558
2	0.046 174	0.080 260	0.104 558
3	0.084 368	0.108 440	0.104 558
4	0.103 549	0.099 330	0.104 558

表 4.11 耦合系统(4.77)在 $\varepsilon=0.012$ 时的最大 Lyapunov 指数的比较

变量个数 l	时间序列长度为 2000 时计算的 λ_1	时间序列长度为 $2000/l$ 计算的 λ_1	λ_1 的理论值
1	0.083 325	0.083 325	0.128 993
2	0.098 919	0.088 860	0.128 993
3	0.106 871	0.109 560	0.128 993
4	0.121 149	0.124 510	0.128 993

表 4.8～表 4.11 的计算结果表明：采用多变量时间序列计算时，随着变量数量的增加，对于时间序列的长度可以明显减少，而计算结果相差不多。例如，采用 4 个变量长为 500 的时间序列和采用 4 个变量长为 2000 的时间序列计算的结果十分接近，而且始终优于单变量、2 个变量和 3 个变量长为 2000 的时间序列的计算结果。因此，在观测时间序列长度有限的情况下，通过多变量时间序列重构从某种程度上可以弥补时间序列长度的不足。

4.5　多变量时间序列相空间重构中噪声的影响

观测或实验获得的时间序列常常含有噪声，下面通过例 4.5 中的耦合系统进行仿真分析，来比较噪声对单变量时间相空间重构和多变量时间序列相空间重构的影响程度。

定义含有测量噪声（假设为高斯噪声）的信号为 y，$y=x+\xi$，x 是观测系统的真实数据，ξ 代表噪声，且 $z=\sigma_\xi^2/\sigma_x^2$，$\sigma^2$ 是相应数据的方差，z 代表噪声级别，下面的仿真试验中分别讨论 $z=2\%$ 和 $z=20\%$ 两种情况。

对例 4.5 中的耦合系统通过四阶 Runge-Kuta 积分法得到的时间序列，图 4.9 和图 4.10 分别是当 $\varepsilon=0.05$ 和 $\varepsilon=0.50$ 时对时间序列 $\{x_1\}$ 加了 2% 噪声后利用改进的最近邻点法计算嵌入维数的过程，图 4.9 和图 4.10 中曲线表示随着搜索全局重构相空间维数的增加，算法中 $E_1(m)$ 逐步收敛的过程。其中浅色虚线和浅色实线分别表示采用没有噪声的 $\{x_1\}$ 和附加 2% 噪声后的 $\{x_1\}$ 进行单变量相空间重构时的情形；深色虚线为附加 2% 噪声后的 $\{x_1\}$ 和没有噪声的 $\{x_2\}$ 进行两变量相空间重构时的情形；深色实线为附加 2% 噪声后的 $\{x_1\}$、没有噪声的 $\{x_2\}$ 和 $\{y_1\}$ 进行三变量相空间重构的情形。

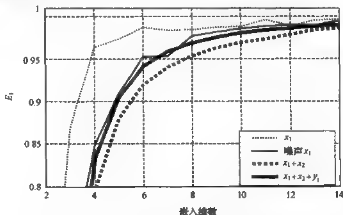


图 4.9　耦合系数 $\varepsilon=0.05$ 噪声级别 $z=2\%$ 时嵌入维数的收敛过程

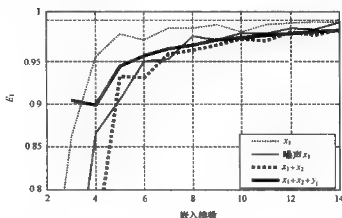


图 4.10 耦合系数 $\varepsilon=0.50$ 噪声级别 $z=2\%$时嵌入维数的收敛过程

图 4.11 和图 4.12 分别是当 $\varepsilon=0.05$ 和 $\varepsilon=0.50$ 时对时间序列$\{x_1\}$加了 20%噪声后，利用改进的最近邻点法计算嵌入维数的过程，图 4.11 和图 4.12 中曲线表示随着搜索全局重构相空间维数的增加，算法中 $E_1(m)$逐步收敛的过程。其中浅色虚线和浅色实线分别表示采用没有噪声的$\{x_1\}$和附加 20%噪声后的$\{x_1\}$进行单变量相空间重构时的情形；深色虚线为附加 20%噪声后的$\{x_1\}$和没有噪声的$\{x_2\}$进行两变量相空间重构时的情形；深色实线为附加 20%噪声后的$\{x_1\}$、没有噪声的$\{x_2\}$和$\{y_1\}$进行三变量相空间重构的情形。

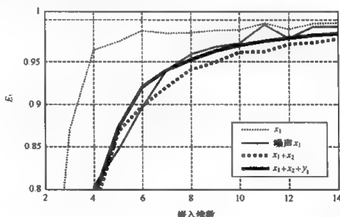


图 4.11 耦合系数 $\varepsilon=0.05$ 噪声级别 $z=20\%$时嵌入维数的收敛过程

根据改进的最近邻点算法求得的嵌入维数的结果见表 4.12。其中第 3 列表示没有噪声的单变量时间序列$\{x_1\}$的重构参数，第 4 列表示附加噪声的单变量时间

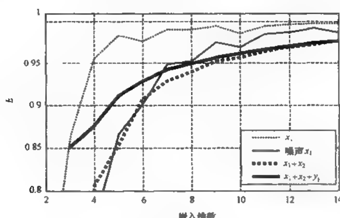


图 4.12　耦合系数 $\varepsilon=0.50$ 噪声级别 $z=20\%$ 时嵌入维数的收敛过程

序列$\{x_1\}$的重构参数，第 5 列表示附加噪声的时间序列$\{x_1\}$、没有噪声的时间序列$\{x_2\}$进行 2 变量重构时的重构参数；第 6 列表示附加噪声的时间序列$\{x_1\}$、没有噪声的时间序列$\{x_2\}$和$\{y_1\}$进行 3 变量重构时的重构参数。

表 4.12　不同噪声级别下相空间重构参数的计算结果

噪声级别	耦合强度	无噪 x_1	有噪 x_1	有噪 x_1 +无噪 x_2	有噪 x_1+无噪 x_2 +无噪 y_1
2%	0.05	6/17	8/11	5/11,2/14	5/11,2/14,1/14
	0.50	7/14	8/16	6/16,1/14	5/16,1/14,1/14
20%	0.05	6/17	13/10	6/10,2/14	6/10,2/14,1/11
	0.50	7/14	12/11	7/10,1/14	8/10,1/14,1/14

从图 4.9～图 4.12 中可以看出，利用没有噪声的测量数据进行嵌入空间的搜索时，算法可以有效地收敛，在 $\varepsilon=0.05$ 时，算法可以给出正确的重构相空间维数 6，在 $\varepsilon=0.50$ 时，由于混沌耦度加强，耦合系统的特性发生变化给出重构维数 7。采用受噪声干扰的$\{x_1\}$进行单变量重构时算法不能在实际嵌入空间维数处收敛，只能勉强收敛在较大的嵌入维数。对比图 4.9 和图 4.11、图 4.10 和图 4.12 中的浅色实线和两个深色线段可以发现，尽管使用了附加噪声的数据，但进行两变量时间序列和三变量时间序列相空间重构时，算法可以给出较好的收敛性。当噪声级别较低时，算法收敛时给出的重构维数与实际重构维数基本一致；当噪声级别较大时，算法虽然收敛得比低噪声级别慢，但仍可以给出可用的重构维数。

对比图 4.9～图 4.12 中的深色的实线(三变量)和虚线(两变量)可以发现，采用三变量时间序列重构可以给出比两变量时间序列重构时更加平稳的收敛变化，

特别是图 4.10 和图 4.11 中可以明显看出这种差异，这种现象可以认为是由于在利用多变量重构时，较多的观测数据可以引入更加丰富的原动力学系统所包含的信息，因此可以反映出更好的特性。

从图 4.9 中可以看出，采用三变量时间序列重构可以比采用两变量时间序列或单变量时间序列重构在更小的全局重构维数处达到收敛，这种现象在其他仿真结果中也存在，这表明采用多变量时间序列重构相空间的优越性。当有噪声干扰时，采用单变量时间序列重构相空间时，算法由于受到噪声干扰无法在实际重构维数处达到收敛，直到较大维数时才能收敛。当噪声级别较小(如 2%)时采用三变量时间序列重构，本算法可以给出与无噪声干扰时的相近结果；当噪声级别较大(如 20%)时采用三变量时间序列重构，算法可以在 8～9 维处收敛，与无噪声干扰时的结果差距不大。

第五章　多变量非线性时间序列预测方法

对某一未知数学模型的确定性系统通过观测或实验手段获得的单变量或多变量时间序列的研究中，一个重要的问题在于能否以及如果根据观测时间序列来推知整个动力系统的变化情况，即非线性时间序列的预测问题。利用单变量或多变量时间序列的相空间重构理论，可以对来自于确定性系统的非线性时间序列包括混沌时间序列进行短期预测。本章以多变量时间序列为研究对象，介绍其预测方法，单变量时间序列可以作为多变量时间序列的特例。

设观测到的 L 个变量非线性时间序列为 $\{x_{l,n}\}_{n=1}^{T}(l=1,2,\cdots,L)$，假设 $T=N+P$，前 N 个数据作为构造模型所需的样本，后 P 个数据作为预测精度的度量。采用延迟重构法，在相空间中重构的状态向量为

$$x_n = \left(x_{1,n}, x_{1,n-\tau_1}, \cdots, x_{1,n-(m_1-1)\tau_1}; x_{2,n}, x_{2,n-\tau_2}, \cdots, x_{2,n-(m_2-1)\tau_2}; \cdots; \right.$$
$$\left. x_{L,n}, x_{L,n-\tau_L}, \cdots, x_{L,n-(m_L-1)\tau_L}\right) \in R^m$$
$$n = N_0, N_0+1, \cdots, N \tag{5.1}$$

式中，$N_0=\max\limits_{1\leq l\leq L}\{(m_l-1)\tau_l+1\}$；$\tau_l(l=1,2,\cdots,L)$和 $m_l(l=1,2,\cdots,L)$分别表示延迟时间间隔和嵌入维数，类似于 F. Takens 在单变量情形下的结论，一般地，如果 m_l 或 $m=\sum\limits_{l=1}^{L} m_l$ 充分大，则 x_n 是原动力系统相应的一条轨道到 R^m 中的嵌入。由此可得到 R^m 上的一个离散的动力系统 $G: R^m \rightarrow R^m$，使得

$$x_{n+1} = G(x_n) \tag{5.2}$$

或者得到多元函数 $G_l: R^m \rightarrow R$，满足

$$x_{l,n+1} = G_l(x_n),\ l = 1,2,\cdots,L \tag{5.3}$$

非线性时间序列预测问题就是根据 $\{x_{l,n}\}_{n=1}^{T}(l=1,2,\cdots,L)$构造 G_l（或 G）的一个近似形式 $\hat{G}_l$（或 $\hat{G}$）。当然，由于系统的混沌性质，$\hat{G}_l$ 或 $\hat{G}$ 应该是非线性的。以下主要讨论对第一个变量 $x_{1,n}$的预测问题，其他变量的预测类似可得。

5.1　多变量非线性时间序列的局域预测法

局域预测法只利用被预测点邻域内其他点的信息来拟合预测函数，不同的预测点处的预测函数是不一样的。常用的局域预测法有局部平均预测法、局部线性预测法、局部多项式预测法等。

5.1.1　局部平均预测法

设时刻 T 的状态向量为

$$\boldsymbol{x}_T=(x_{1,T},x_{1,T-\tau_1},\cdots,x_{1,T-(m_1-1)\tau_1};x_{2,T},x_{2,T-\tau_2},\cdots,x_{2,T-(m_2-1)\tau_2};\\ \cdots;x_{L,T},x_{L,T-\tau_L},\cdots,x_{L,T-(m_L-1)\tau_L})\tag{5.4}$$

找 $\boldsymbol{x}_T$ 的 K 个最近邻点 $\boldsymbol{x}_{e_1},\boldsymbol{x}_{e_2},\cdots,\boldsymbol{x}_{e_K}$，如果系统是确定性的，则当 $\boldsymbol{x}_T$ 与 $\boldsymbol{x}_{e_k}(k=1,2,\cdots,K)$靠近时，$\boldsymbol{x}_{T+1}$也与 $\boldsymbol{x}_{e_k+1}(k-1,2,\cdots,K)$靠近。因此，以 $x_{1,e_1+1},x_{1,e_2+1},\cdots,x_{1,e_k+1}$的平均作为 $x_{1,T+1}$的预测值，即

$$\hat{x}_{1,T+1}=\frac{1}{K}\sum_{k=1}^{K}x_{1,e_k+1}\tag{5.5}$$

这种预测方法称为局部平均预测法。

单变量情形下的局部平均预测法可见文献[74]，作为特例，也可取 $K=1$，此时称为零级近似预测[120]。

局部平均预测法在找到预测点的邻域后，没有考虑邻域中的各点到预测点的空间距离对预测的影响，而相空间中的各点到预测点的空间距离是一个非常重要的参数，预测的准确性往往取决于与预测点空间距离最近的那几个点。因此，可以将到中心点的空间距离作为一个拟合参数引入预测过程，在一定程度上可以提高预测精度，并有一定的降噪能力。基于此，参照文献[121]在单变量时间序列情形下的改进，可做如下修改：

$$\hat{x}_{1,T+1}=\frac{\sum\limits_{k=1}^{K}x_{1,e_{k+1}}\exp[-\omega(d_k-d_{\min})]}{\sum\limits_{k=1}^{K}\exp[-\omega(d_k-d_{\min})]}\tag{5.6}$$

式中，$d_k=\|\boldsymbol{x}_T-\boldsymbol{x}_{e_k}\|$，$k=1,2,\cdots,K$，$d_{\min}=\min\limits_{k=1}^{K}d_k$，$\omega$ 为参数，一般取 $\omega\geqslant1$。

上述模型中，参数 ω 的选择有主观性，为此参照文献[122]在单变量时间序列情形下的改进，同样可采用以下改进的预测模型：

$$\hat{x}_{1,T+1}=\frac{\sum\limits_{k=1}^{K}x_{1,e_{k+1}}\dfrac{d_{\max}-d_k}{d_{\max}-d_{\min}}}{\sum\limits_{k=1}^{K}\dfrac{d_{\max}-d_k}{d_{\max}-d_{\min}}}\tag{5.7}$$

其中 $d_{\max}=\max\limits_{k=1}^{K}d_k$。

5.1.2　局部线性预测法

设时刻 T 的状态向量如式(5.4)所示，则对 p 步之后的值 $x_{1,T+p}$进行预测，预

测模型为

$$\begin{aligned}\hat{x}_{1,T+p}=\hat{G}_l(x_T)=&c_0+c_1x_{1,T}+c_2x_{1,T-\tau_1}+\cdots+c_{m_1}x_{1,T-(m_1-1)\tau_1}\\&+c_{m_1+1}x_{2,T}+c_{m_1+2}x_{2,T-\tau_2}+\cdots+c_{m_1+m_2}x_{2,T-(m_2-1)\tau_2}\\&+\cdots\cdots\cdots\\&+c_{m_1+\cdots+m_{L-1}+1}x_{L,T}+c_{m_1+\cdots+m_{L-1}+2}x_{L,T-\tau_L}+\cdots+c_{m_1+\cdots+m_L}x_{L,T-(m_L-1)\tau_L}+e\end{aligned} \tag{5.8}$$

式中，$m=m_1+m_2+\cdots+m_L$；$c_0,c_1,c_2,\cdots,c_m$ 为待定系数；e 为随机误差，具有期望 $E(e)=0$ 和方差 $\mathrm{Var}(e)=\sigma^2$。

为了估计这些系数，计算目标点 $\boldsymbol{x}_T$ 与各个重构向量间的欧氏距离，得到目标点 $\boldsymbol{x}_T$ 的 K 个最近邻点 $\boldsymbol{x}_{a_1},\boldsymbol{x}_{a_2},\cdots,\boldsymbol{x}_{a_K}$，即

$$\begin{aligned}\boldsymbol{x}_{a_k}=\big(&x_{1,a_k},x_{1,a_k-\tau_1},\cdots,x_{1,a_k-(m_1-1)\tau_1};x_{2,a_k},x_{2,a_k-\tau_2},\cdots,x_{2,a_k-(m_2-1)\tau_2};\\&\cdots\cdots;x_{L,a_k},x_{L,a_k-\tau_L},\cdots,x_{L,a_k-(m_L-1)\tau_L}\big)\end{aligned} \tag{5.9}$$

则当 $\boldsymbol{x}_T$ 与 $\boldsymbol{x}_{a_k}$，$k=1,2,\cdots,K$ 靠近时，$\boldsymbol{x}_{T+p}$ 也与 $\boldsymbol{x}_{a_k+p}$，$k=1,2,\cdots,K$ 靠近。于是问题转化为求 $c_0,c_1,c_2,\cdots,c_m$，使

$$\sum_{k=1}^{K}\left|x_{1,a_k+p}-\hat{x}_{1,T+p}\right|^2=\sum_{k=1}^{K}\left|x_{1,a_k+p}-\hat{G}_l(\boldsymbol{x}_T)\right|^2 \tag{5.10}$$

最小。若记

$$\boldsymbol{y}=\left[x_{1,a_1+p},x_{1,a_2+p},\cdots,x_{1,a_K+p}\right]^{\mathrm{T}}\in R^K \tag{5.11}$$

$$\boldsymbol{c}=\left[c_0,c_1,c_2,\cdots,c_m\right]^{\mathrm{T}}\in R^{m+1} \tag{5.12}$$

$$\boldsymbol{X}=\begin{bmatrix}1 & x_{1,a_1} & x_{1,a_1-\tau_1} & \cdots & x_{L,a_1-(m_L-1)\tau_L}\\ 1 & x_{1,a_2} & x_{1,a_2-\tau_1} & \cdots & x_{L,a_2-(m_L-1)\tau_L}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & x_{1,a_K} & x_{1,a_K-\tau_1} & \cdots & x_{L,a_K-(m_L-1)\tau_L}\end{bmatrix}\in R^{K\times(m+1)} \tag{5.13}$$

则利用最小二乘原理，$c_0,c_1,c_2,\cdots,c_m$ 满足下面的矩阵方程：

$$\boldsymbol{y}=\boldsymbol{X}\boldsymbol{c} \tag{5.14}$$

如果矩阵 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 可逆，则有

$$\boldsymbol{c}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \tag{5.15}$$

从而可由公式(5.8)计算 $x_{1,T+p}$ 的预测值 $\hat{x}_{1,T+p}$。

预测步数 $p=1$ 时称为一步预测，$p>1$ 时称为多步预测。

在上述线性回归模型中一般要求 $K>m$，但理论上对 $K\leqslant m$ 的情形也同样适用。单变量时间序列情形下这一方法可见文献[123]。关于参数 K 的选取，单变量时间序列情形下，文献[124]给出了下面的方法：

步骤 1：取 $K=2m+1,2m+2,\cdots,2m+10$。

步骤 2：对上述每一 K，计算

$$H = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} \tag{5.16}$$

$$D(K) = \mathrm{tr}(H) \tag{5.17}$$

$$\sigma^2(K) = \frac{[Y - YX^{\mathrm{T}}(XX^{\mathrm{T}})^{-1}X][Y - YX^{\mathrm{T}}(XX^{\mathrm{T}})^{-1}X]^{\mathrm{T}}}{K - D(K)} \tag{5.18}$$

步骤 3：选择使 $\sigma^2(K)$ 最小的 K，记为 K^*。

步骤 4：利用 K^* 构造局部预测模型，即计算 C，然后进一步计算 $x_{1,T+p}$ 的预测值 $\hat{x}_{1,T+p}$。

5.1.3　局部多项式预测法

单变量时间序列局部多项式预测法是由 J. D. Farmer 和 J. J. Sidorowich 给出的[125]，选用的映射是高阶多项式，其系数用最近邻点拟合，文献[126]给出了确定多项式阶数的方法，用这种方法能够得到比局部平均预测法和局部线性预测法更好的预测结果[127]。既然多变量的局部平均预测法和局部线性预测法均比单变量情形的预测精度高，因此，基于多变量相空间的重构理论，我们把单变量混沌时间序列的局部多项式预测法推广到多变量的情形，给出了多变量混沌时间序列的局部多项式法预测模型[128]。

设时刻 T 的状态向量如式(5.4)所示，对 p 步之后的值 $x_{1,T+p}$ 进行预测，于是 $\boldsymbol{x}_{T+p}$ 的第一个分量 $x_{1,T+p}$ 的预测值 $\hat{x}_{1,T+p}$ 由 s 阶多项式得到，即

$$\begin{aligned}
\hat{x}_{1,T+p} &= \hat{G}_1(\boldsymbol{x}_T) \\
&= c_0 + \sum_{k_1=0}^{m_1-1} c_{1k_1}^{(1)} x_{1,T-k_1\tau_1} + \cdots + \sum_{k_s=k_{s-1},\cdots,k_2=k_1,k_1=0}^{m_1-1} c_{sk_1k_2\cdots k_s}^{(1)} x_{1,T-k_1\tau_1}\cdots x_{1,T-k_s\tau_1} \\
&\quad + \sum_{k_1=0}^{m_2-1} c_{1k_1}^{(2)} x_{2,T-k_1\tau_2} + \cdots + \sum_{k_s=k_{s-1},\cdots,k_2=k_1,k_1=0}^{m_2-1} c_{sk_1k_2\cdots k_s}^{(2)} x_{2,T-k_1\tau_2}\cdots x_{2,T-k_s\tau_2} \\
&\quad + \cdots\cdots\cdots\cdots \\
&= \sum_{k_1=0}^{m_L-1} c_{1k_1}^{(L)} x_{L,T-k_1\tau_L} + \cdots + \sum_{k_s=k_{s-1},\cdots,k_2=k_1,k_1=0}^{m_L-1} c_{sk_1k_2\cdots k_s}^{(L)} x_{L,T-k_1\tau_L}\cdots x_{L,T-k_s\tau_L}
\end{aligned} \tag{5.19}$$

找 $\boldsymbol{x}_T$ 的 K 个最近邻点 $\boldsymbol{x}_{a_1},\boldsymbol{x}_{a_2},\cdots,\boldsymbol{x}_{a_K}$，如果系统是确定的，则当 $\boldsymbol{x}_T$ 靠近 $\boldsymbol{x}_{a_k}$ 时，$\boldsymbol{x}_{T+p}$ 也靠近 $\boldsymbol{x}_{a_k+p}$。

于是问题转化为求系数 c_0，$c_{10}^{(1)}$，$\cdots$，$c_{1(m_1-1)}^{(1)}$，$\cdots$，$c_{10}^{(L)}$，$\cdots$，$c_{1(m_L-1)}^{(L)}$，$\cdots$，$c_{s(m_L-1)\cdots(m_L-1)}^{(L)}$，使

$$\sum_{k=1}^{K} \left| x_{1,a_k+p} - \hat{x}_{1,T+p} \right|^2 = \sum_{k=1}^{K} \left| x_{1,a_k+p} - \hat{G}_1(\boldsymbol{x}_T) \right|^2 \tag{5.20}$$

最小。若记

$$\boldsymbol{y} = \left[x_{1,e_1+p}, x_{1,e_2+p}, \cdots, x_{1,e_K+p}\right]^{\mathrm{T}} \in R^K \tag{5.21}$$

$$\boldsymbol{c} = \left[c_0, c_{10}^{(1)}, \cdots, c_{1(m_1-1)}^{(1)}, \cdots, c_{10}^{(L)}, \cdots, c_{1(m_L-1)}^{(L)}, \cdots, c_{s(m_L-1)\cdots(m_L-1)}^{(L)}\right]^{\mathrm{T}} \in R^J \tag{5.22}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{1,e_1} & \cdots & x_{1,e_1-(m_1-1)\tau_1} & x_{1,e_1}^2 & \cdots & x_{L,e_1} & \cdots & x_{L,e_1-(m_L-1)\tau_L}^s \\ 1 & x_{1,e_2} & \cdots & x_{1,e_2-(m_1-1)\tau_1} & x_{1,e_2}^2 & \cdots & x_{L,e_2} & \cdots & x_{L,e_2-(m_L-1)\tau_L}^s \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1,e_K} & \cdots & x_{1,e_K-(m_1-1)\tau_1} & x_{1,e_K}^2 & \cdots & x_{L,e_K} & \cdots & x_{L,e_K-(m_L-1)\tau_L}^s \end{bmatrix} \in R^{K\times J} \tag{5.23}$$

式中待定系数个数 J 为

$$J = \sum_{l=1}^{L} C_{m_l+s}^{s} - L + 1 \tag{5.24}$$

则利用最小二乘原理，系数 $c_0, c_{10}^{(1)}, \cdots, c_{1(m_1-1)}^{(1)}, \cdots, c_{10}^{(L)}, \cdots, c_{1(m_L-1)}^{L}, \cdots, c_{s(m_L-1)\cdots(m_L-1)}^{(L)}$ 满足下面的矩阵方程

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{c} \tag{5.25}$$

为了获得稳定的解，一般要求 $K \geqslant J$。如果矩阵 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 可逆，则有

$$\boldsymbol{c} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \tag{5.26}$$

从而可由式(5.19)计算 $x_{1,T+p}$ 的预测值 $\hat{x}_{1,T+p}$。

显然，局部多项式预测法推广了局部线性预测法，同局部线性预测法一样，预测步数 $p=1$ 时称为一步预测，$p>1$ 时称为多步预测。

单变量时间序列情形下的局域预测方法还有基于最大 Lyapunov 指数的预测方法[129, 130]、线性内插方法[131, 132]、局域超平面近似法[131, 133]。

局域预测法柔性较好，拟合速度较快且精度较高，但需要实时地拟合系数。局域预测方法的预测函数虽然在每一步都可能是线性的或简单的多项式，但从整体上看却是非线性的。

5.2　多变量非线性时间序列的全域预测法

所谓全域预测法是将轨线中的所有点作为拟合对象，寻找合适的函数 G_l。由于实际问题中数据总是有限的，因而不可能求出真正的 G_l，只能通过有限数据构造 $\hat{G}_l$，使

$$\sum_{n}\left[x_{l,n+1} - G_l(\boldsymbol{x}_n)\right]^2 \tag{5.27}$$

最小，从而可得全局模型

$$x_{l,n+1} = \hat{G}_l(\boldsymbol{x}_n) \tag{5.28}$$

再用它的迭代来预测未来每一步。

全局法在概念上很清楚，但实现并不容易，尤其当映射 G_l 很复杂时，所以预测时必须找出 G_l 的一个非线性表达式。

5.2.1　多项式逼近预测法

通常多项式是非线性表达式的一个很好近似，多项式逼近预测法就是从标准函数基中选取多项式 G_l，G_l 的系数由最小二乘法确定，当系数个数和多项式的阶数增加时，所拟合的函数 G_l 会越来越接近真实的动力系统。

取 $G_l(\boldsymbol{x}_n)$ 为 $x_{1,n},x_{1,n-\tau_1},\cdots,x_{1,n-(m_1-1)\tau_1};\cdots;x_{L,n},x_{L,n-\tau_L},\cdots,x_{L,n-(m_L-1)\tau_L}$ 共 m 个变量的 s 次多项式，即

$$x_{l,n+1}=G_l(\boldsymbol{x}_n)$$

$$=\sum_{(i_1,\cdots,i_m)\in S}a_{i_1\cdots i_m}^{(l)}x_{1,n}^{i_1}x_{1,n-\tau_1}^{i_2}\cdots x_{1,n-(m_1-1)\tau_1}^{i_{m_1}}\cdots x_{L,n}^{i_{m_1+\cdots+m_{L-1}+1}}\cdots x_{L,n-(m_L-1)\tau_L}^{i_m} \tag{5.29}$$

其中

$$S=\{(i_1,i_2,\cdots,i_m)\,|\,i_1+i_2+\cdots+i_m\leqslant s,\ i_1,i_2,\cdots,i_m \text{ 为大于等于零的整数}\} \tag{5.30}$$

当 s 和 m 较大时，待定系数 $a_{i_1\cdots i_m}^{(l)}$ 较多，利用最小二乘法确定起来非常复杂。

多项式逼近预测法的缺点是当多项式的阶数很大时，要确定大量系数，这往往会超过计算机的承受能力而使计算无法进行下去。另外，这时也要求有足够的样本点。

5.2.2　神经网络预测法

用于预测研究的神经网络模型主要有多层前馈型神经网络，一般的前馈型神经网络具有逼近复杂非线性函数的能力，通常可以根据所选取的样本数据对网络进行训练，在训练过程中使预测的均方差达到最小，从而建立相应的预测模型。由于一个 3 层前馈网络有可能逼近任意的复杂连续函数，所以人工神经网络也是逼近非线性函数的重要预测工具。

已知输入 m 维向量

$$\boldsymbol{x}_n=(x_{1,n},x_{1,n-\tau_1},\cdots,x_{1,n-(m_1-1)\tau_1};\cdots;x_{L,n},x_{L,n-\tau_L},\cdots,x_{L,n-(m_L-1)\tau_L}) \tag{5.31}$$

对应的 L 维输出向量为

$$\boldsymbol{y}_n=(x_{1,n+1},x_{2,n+1},\cdots,x_{L,n+1}) \tag{5.32}$$

其中 $n=N_0,N_0+1,\cdots,N$，这些向量称为样本，通过训练（或学习），找向量函数 $\hat{G}$，使

$$\boldsymbol{y}_n=\hat{G}(\boldsymbol{x}_n),\ n=N_0,N_0+1,\cdots,N \tag{5.33}$$

这时 BP 网络结构如图 5.1 所示。

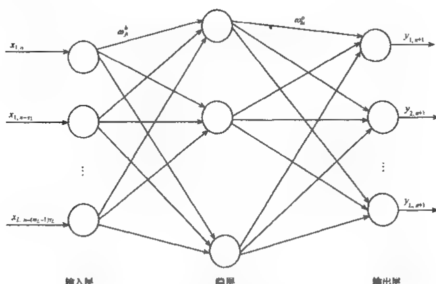


图 5.1　多变量时间序列预测的 BP 网络图

上述 BP 网络结构图中，有 m 个输入节点，L 个输出节点，隐层有 q 个节点，ω^h_{ji} 是第 i 个输入节点和第 j 个隐含节点之间的连接权值，ω^o_{kj}是第 j 个隐含层和第 k 个输出节点之间的连接权值。则神经网络的输入输出关系为

$$x_{k,n+1} = f\Big(\sum_{j=1}^{q}\omega^o_{kj}O_j\Big) = f\Big(\sum_{j=1}^{q}\omega^o_{kj}f\Big(\sum_{i=1}^{m}x_i\omega^h_{ji}\Big)\Big),\ k=1,2,\cdots,L \tag{5.34}$$

其中输入 $x_1 = x_{1,n}, x_2 = x_{1,n-\tau_1}, \cdots, x_{m_1} = x_{1,n-(m_1-1)\tau_1}, \cdots, x_{m_1+\cdots+m_{L-1}+1} = x_{L,n}$, $x_{m_1+\cdots+m_{L-1}+2} = x_{L,n-\tau_L}, \cdots, x_m = x_{L,n-(m_L-1)\tau_L}$，激发函数取 $f(s)=\dfrac{1}{1+\mathrm{e}^{-s}}$。

通过 $n=N_0, N_0+1, \cdots, N$ 个样本的训练后，函数 G 就由网络中的权值 ω^h_{ji}和 ω^o_{kj}确定。由上述网络可以计算出 $\hat{x}_{1,T+1}, \hat{X}_{1,T+2}, \cdots, \hat{x}_{1,T+P}$，它们分别是 $x_{1,T+1}$, $x_{1,T+2}, \cdots, x_{1,T+P}$的预测值。

单变量时间序列情形下基于神经网络的预测方法的应用可见文献[134~141]，利用神经网络进行预测的其他方法还有小波神经网络模型[142,143]、模糊神经网络预测模型[144~146]等。

5.2.3　径向基函数预测法

BP 网络为全局逼近网络，在用于函数逼近时，权值的调整采用梯度下降法，存在收敛速度慢、局部极值等缺点。径向基函数神经网络为局部逼近网络，对于每个输入输出数据时，只有少量的权值需要调整，具有学习速度快的优点，因此径向基

函数神经网络在逼近能力、分类能力和学习速度等方面均优于 BP 网络。

径向基函数预测模型为(以预测第一个变量为例)

$$x_{1,n+1} = \hat{G}_1(x_n) = \sum_{j=1}^{N_c} \lambda_j \varphi_j(\| x_n - c_j \|), \ n = N_0, N_0+1, \cdots, N \quad (5.35)$$

其中 $\varphi_j(r): R^+ \to R$ 是径向基函数,一般取为

$$\varphi_j(r) = \exp\left(-\frac{r^2}{2\sigma_j^2}\right) \quad (5.36)$$

式中,σ_j 为常数,称为宽度;$c_j(j=1,2,\cdots,N_c)$为经向基函数的中心;$\lambda_j(j=1,2,\cdots,N_c)$是待定常数。

径向基函数预测模型中待确定的参数有两类:径向基函数的中心点 $c_j(j=1,2,\cdots,N_c)$和宽度 $\sigma_j(j=1,2,\cdots,N_c)$;待定常数 $\lambda_j(j=1,2,\cdots,N_c)$。有不少方法来确定这两类参数。

1. 径向基函数的中心点和宽度的选择

基本思想:利用输入样本来计算 N_c 个 $c_j(j=1,2,\cdots,N_c)$,目的是使 c_j 尽可能均匀地对输入数据抽样,使 c_j 能够反映输入向量在输入空间的分布情况,在输入向量集中的地方选中的 c_j 也多,因此径向基函数中心的选择就是对输入特性认识和分类的过程。径向基函数的中心点和宽度的选择方法常用的有 K-均值聚类算法[147]、硬 C 均值算法[148,149]和 Kohonen 自组织神经网络法[150~152]。

(1) K-均值聚类算法

这是一种无监督的学习算法,具体步骤如下:

步骤 1:初始化聚类中心 $c_j(j=1,2,\cdots,N_c)$。一般是从输入样本$\{x_n\}$ $(n=N_0, N_0+1,\cdots,N)$中选择一个样本作为聚类中心;

步骤 2:将输入样本按照最邻近规则分组,即将输入样本$\{x_n\}$ $(n=N_0, N+1,\cdots,N)$分配给中心为 $c_j(j=1,2,\cdots,N_c)$的输入样本聚类集合 $\theta_j(j=1,2,\cdots,N_c)$,亦即 $x_n \in \theta_j$ 且满足

$$d_j = \min_n \| x_n - c_j \|, \ n = N_0, N_0+1, \cdots, N; \ j = 1,2,\cdots,N_c \quad (5.37)$$

步骤 3:计算 θ_j 中样本的平均值即聚类中心

$$c_j = \frac{1}{M_j} \sum_{x_n \in \theta_j} x_n \quad (5.38)$$

式中,M_j 为 θ_j 的输入样本数。

步骤 4:重复上述步骤,直到聚类中心不再变化为止。

聚类中心确定后,可由下式计算宽度 $\sigma_j(j=1,2,\cdots,N_c)$,即

$$\sigma_j^2 = \frac{1}{M_j} \sum_{x_n \in \theta_j} \| x_n - c_j \|^2 \quad (5.39)$$

(2) 硬 C 均值算法

设输入样本 $\{x_n\}$ $(n=N_0,N_0+1,\cdots,N)$ 可分成 N_c 类，用一个 $N_c\times(N-N_0)$ 矩阵 U 来表示分类结果，U 中的元素 u_{jn} 为

$$u_{jn}=\begin{cases}1,\text{当 } x_n\in A_j \text{ 时}\\0,\text{当 } x_n\notin A_j \text{ 时}\end{cases}\tag{5.40}$$

式中，A_j 表示第 j 类。

算法的具体步骤如下：

步骤 1：确定聚类类别数，$2\leqslant N_c\leqslant N-N_0$。

步骤 2：预置允许误差 ε。

步骤 3：任意指定初始分类矩阵 U^b，$b=0$。

步骤 4：根据 U^b 和下式计算 N_c 个中心 $c_j(j=1,2,\cdots,N_c)$：

$$c_j=\frac{1}{\sum\limits_{n=1}^{N-N_0}u_{jn}}\sum_{n=1}^{N-N_0}u_{jn}x_n\tag{5.41}$$

步骤 5：按如下方式更新 U^b 为 U^{b+1}：

$$u_{jn}^{b+1}=\begin{cases}1,\text{当 } d_{jn}^b=\min\limits_{1\leqslant i\leqslant N_c}\{d_{in}^b\}\text{ 时}\\0,\text{其他}\end{cases}\qquad j=1,2,\cdots,N_c;\ n=N_0+1,\cdots,N\tag{5.42}$$

式中，$d_{jn}=\|x_n-c_j\|$，即第 n 个样本 x_n 到第 j 个中心 c_j 之间的距离。

步骤 6：以一个合适的矩阵范数比较 U^b 和 U^{b+1}，若 $\|U^b-U^{b+1}\|<\varepsilon$，则停止；否则置 $b=b+1$，返回到步骤 4。

此时，宽度 $\sigma_j(j=1,2,\cdots,N_c)$ 由式(5.43)得到

$$\sigma_j=\frac{1}{\sum\limits_{n=1}^{N-N_0}u_{jn}}\sum_{n=1}^{N-N_0}u_{jn}d_{jn}\tag{5.43}$$

(3) Kohonen 自组织神经网络法

基本思想：将输入样本 $\{x_n\}$ $(n=N_0,N_0+1,\cdots,N)$ 选择为用于构造模型的学习点集，选择输入层由 m 个神经元组成，分别对应 x_n 的 m 个分量，输出层直接取成 N_c 个神经元，表示希望产生的中心个数，每个输出神经元都与 m 个输入神经元直接连接，其连接权为 $\omega_{ij}(n)$ $(i=1,2,\cdots,m;j=1,2,\cdots,N_c)$，采用常规的学习算法，调整 $\omega_{ij}(n)$ 直至其收敛，把每一输出神经元对应的最终稳定的权向量 $\boldsymbol{\omega}_j(j=1,2,\cdots,N_c)$ 作为预测模型的中心点 $c_j(j=1,2,\cdots,N_c)$。

2. 待定常数 λ_j 的求法

求待定常数 λ_j 的方法有矩阵算法[153,154]和迭代算法。

(1) 矩阵算法

在确定了径向基函数的中心点和宽度后，求 $\lambda_j(j=1,2,\cdots,N_c)$，使

$$\sum_{n=N_0+1}^{N}\left[x_{1,n+1}-\sum_{j=1}^{N_c}\lambda_j\varphi_j(\|\boldsymbol{x}_n-\boldsymbol{c}_j\|)\right]^2 \tag{5.44}$$

最小。由最小二乘法可知，若记

$$\boldsymbol{y}=[x_{1,N_0+2},x_{1,N_0+3},\cdots,x_{1,N+1}]^{\mathrm{T}}\in R^{N-N_0} \tag{5.45}$$

$$\boldsymbol{\lambda}=[\lambda_1,\lambda_2,\cdots,\lambda_{N_c}]^{\mathrm{T}}\in R^{N_c} \tag{5.46}$$

$$\boldsymbol{X}=[a_{ij}]\in R^{(N-N_0)\times N_c} \tag{5.47}$$

其中

$$a_{ij}=\varphi_j(\|\boldsymbol{x}_n-\boldsymbol{c}_j\|)\qquad i=N_0+1,N_0+2,\cdots,N;\ j=1,2,\cdots,N_c \tag{5.48}$$

则可从以下矩阵方程求 $\lambda_j(j=1,2,\cdots,N_c)$，即

$$\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\lambda} \tag{5.49}$$

因此当 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 可逆时，有

$$\boldsymbol{\lambda}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \tag{5.50}$$

M. Casdagli[153]在最早把径向基函数用于混沌时间序列的预测时，把中心点 $\boldsymbol{c}_j$ $(j=1,2,\cdots,N_c)$取为训练集中的所有点，即 $\boldsymbol{c}_j=\boldsymbol{x}_j(j=N_0+1,N_0+2,\cdots,N)$，此时，矩阵 $\boldsymbol{A}$ 为方阵，因此得

$$\boldsymbol{\lambda}=\boldsymbol{X}^{-1}\boldsymbol{y} \tag{5.51}$$

这样做的缺点是在每一个中心点都精确地拟合了，因此对一些含噪声的数据，往往会导致拟合过度，不能很好地体现动力系统内在的规律性。另外，对有些数据，$\boldsymbol{A}$ 往往会呈现奇异性，从而无法求出 λ。

(2) 迭代算法

设输入样本为 $\boldsymbol{x}_n(n=N_0,N_0+1,\cdots,N)$，对应的输出值为 $x_{1,n+1}(n=N_0+1,N_0+2,\cdots,N)$，使能量函数

$$E=\frac{1}{2}\sum_{n=N_0+1}^{N}\left[x_{1,n+1}-\sum_{j=1}^{N_c}\lambda_j\varphi_j(\|\boldsymbol{x}_n-\boldsymbol{c}_j\|)\right]^2 \tag{5.52}$$

最小，则迭代公式为

$$\lambda_j^{k+1}=\lambda_j^k+\sum_{n=N_0+1}^{N}\eta\left[x_{1,n+1}-\sum_{j=1}^{N_c}\lambda_j\varphi_j(\|\boldsymbol{x}_n-\boldsymbol{c}_j\|)\right]\varphi_j(\|\boldsymbol{x}_n-\boldsymbol{c}_j\|) \tag{5.53}$$

式中，η 为学习率。

在确定了所有参数后，则后 P 个数据 $x_{1,N+p}(p=1,2,\cdots,P)$的预测值 $\hat{x}_{1,N+p}$ $(p=1,2,\cdots,P)$为

$$\hat{x}_{1,N+p} = \sum_{j=1}^{N_c} \lambda_j \varphi_j(\| x_{T+p-1} - c_j \|),\ p = 1,2,\cdots,P \tag{5.54}$$

而模型的拟合精度可由下面的拟合均方误差表示，即

$$\sigma_f^2 = \frac{1}{N - N_0} \sum_{n=N_0+1}^{N} (x_{1,n+1} - \hat{G}_1(x_n))^2 \tag{5.55}$$

5.3　各种预测方法的预测效果对比分析

针对前面提出的各种局域和全局预测方法，它们的预测模型是否精确、预测的效果如何等是需要进一步探讨的问题，为此将引入预测效果的评价指标，并通过已知模型进行仿真分析。

5.3.1　预测效果评价

为了检验预测的精确性，可以通过比较预测值 $\hat{x}_{1,T+1}$ 与观测值 $x_{1,T+1}$，定义预测误差为 $\hat{x}_{1,T+1}$ 与 $x_{1,T+1}$ 的差。一次预测可能较好或较差，偶然性较大，为了克服这种偶然性，可以取多个点的预测误差的平均。

设 $x_{1,T+1},x_{1,T+2},\cdots,x_{1,T+P}$ 的预测值为 $\hat{x}_{1,T+1},\hat{x}_{1,T+2},\cdots,\hat{x}_{1,T+P}$，定义均方根误差(root mean squared error, RMSE)为

$$\mathrm{RMSE} = \sqrt{\frac{1}{P} \sum_{p=1}^{P} (x_{1,T+p} - \hat{x}_{1,T+p})^2} \tag{5.56}$$

或均方误差(mean squared error, MSE)为

$$\mathrm{MSE} = \frac{1}{P} \sum_{p=1}^{P} (x_{1,T+p} - \hat{x}_{1,T+p})^2 \tag{5.57}$$

如果均方根误差 RMSE(或均方误差 MSE)非常大，意味着预测不好，因此时间序列不是来自于确定性系统；反之，如果均方根误差 RMSE(或均方误差 MSE)较小，说明时间序列来自于确定性系统。衡量均方根误差 RMSE(或均方误差 MSE)的大小，仅仅从均方根误差 RMSE(或均方误差 MSE)的数据本身无法确定，因为均方根误差 RMSE(或均方误差 MSE)与观测数据 $\{x_{l,n}\}$ 的数值大小有关，因此在不同问题中的数量级是不一样的。为了解决这一问题，考虑如果每一时刻 $T+p, p=1,2,\cdots,P$，$x_{1,T+1},x_{1,T+2},\cdots,x_{1,T+P}$ 都取相同的预测值 $\hat{x}_1^*$，使平均预测误差 $\varepsilon^* = \frac{1}{P}\sum_{p=1}^{P}(x_{1,T+p} - \hat{x}_1^*)^2$ 最小，则由最小二乘法可知

$$\hat{x}^* = \frac{1}{P} \sum_{p=1}^{P} x_{1,T+p} \tag{5.58}$$

即这样的预测是后 P 长时间序列的均值，因此此时的平均预测误差 ε^* 就是后 P

长时间序列的方差σ^2，由于这种预测是比较差的，因此可用均方根误差RMSE与方差σ进行相对比较，即定义正规化均方根误差(normalized root mean squared error，NRMSE)为

$$\text{NRMSE}=\frac{\text{RMSE}}{\sigma} \tag{5.59}$$

或用均方误差MSE与$\varepsilon^*=\sigma^2$进行相对比较，即定义正规化均方误差(normalized mean squared error，NMSE)为

$$\text{NMSE}=\frac{\text{MSE}}{\sigma^2} \tag{5.60}$$

式中，σ^2是$\{x_{1,T+p}\}_{p=1}^{P}$的方差。

若正规化均方根误差NRMSE(或正规化均方误差NMSE)接近于1，则表明均方根误差RMSE(或均方误差MSE)较大；若正规化均方根误差NRMSE(或正规化均方误差NMSE)接近于0，则表明均方根误差RMSE(或均方误差MSE)较小[74]。

还可以用下面的相关系数来检验预测的效果，定义相关系数为

$$r=\frac{\sum_{p=1}^{P}x_{1,N+p}\hat{x}_{1,N+P}-\sum_{p=1}^{P}x_{1,N+p}\sum_{p=1}^{P}\hat{x}_{1,N+p}}{\sqrt{\sum_{p=1}^{P}\left(x_{1,N+p}-\sum_{p=1}^{P}x_{1,T+p}\right)^2}\cdot\sqrt{\sum_{p=1}^{P}\left(\hat{x}_{1,N+p}-\sum_{p=1}^{P}\hat{x}_{1,T+p}\right)^2}} \tag{5.61}$$

则$0\leqslant r\leqslant 1$，r越接近1表示预测效果越好。

5.3.2 仿真比较

为了比较各种预测方法之间的预测效果、同类预测方法单变量时间序列和多变量时间序列预测方法之间的差异以及时间序列长度对预测效果的影响等，下面通过一个仿真例子来详细分析。

考虑Lorenz系统

$$\begin{cases}\dfrac{\mathrm{d}x_1}{\mathrm{d}t}=\sigma(x_2-x_1)\\[2mm]\dfrac{\mathrm{d}x_2}{\mathrm{d}t}=x_1(r-x_3)-x_2\\[2mm]\dfrac{\mathrm{d}x_3}{\mathrm{d}t}=x_1x_2-bx_3\end{cases} \tag{5.62}$$

其中$\sigma=10$，$r=28$，$b-\frac{8}{3}$。取$x_{1,0}=15.34$，$x_{2,0}=13.68$，$x_{3,0}=37.91$，积分步长$h=0.04$。利用四阶Runge Kuta积分法计算出两个包含12 100个数据的仿真时间序列x_1和x_2。为了减小瞬态的影响，去掉前10 000个数据，剩下的2100个数据中，

前 2000 个数据用来训练样本，后 100 个数据作为预测集。

对单变量时间序列$\{x_{1,n}\}_{n=1}^{2000}$，利用互信息最小法计算得最佳延迟时间 $\tau=4$，如图 5.2 所示；利用虚假最近邻点法得最小嵌入维数取 $m=3$，如图 5.3 所示。因此，对单变量时间序列$\{x_{1,n}\}_{n=1}^{2000}$，做以下相空间重构

$$x_n=(x_{1,n},x_{1,n-4},x_{1,n-8}) \tag{5.63}$$

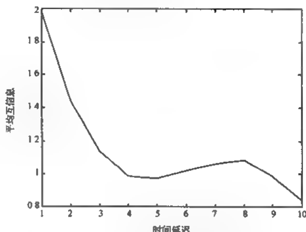


图 5.2　利用平均互信息法求延迟时间

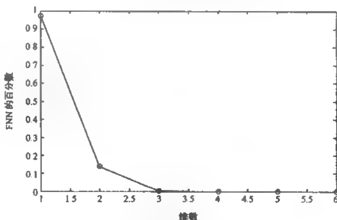


图 5.3　利用虚假最近邻点法求嵌入维数

对两变量时间序列$\{x_{1,n},x_{2,n}\}_{n=1}^{2000}$，由 4.3.4 节的计算知，$\tau_1=\tau_2=4$，$m_1=2$，$m_2=1$，因此，作以下相空间重构

$$x_n=(x_{1,n},x_{1,n-4},x_{2,n}) \tag{5.64}$$

利用前面的局部平均预测法、局部线性预测法、神经网络预测法和径向基函数

预测法，利用式(5.63)和式(5.64)的单变量和多变量重构，分别预测变量 $x_{1,n}$ 的最后 100 个值，即预测 $\{x_{1,2000+p}\}_{p=1}^{100}$，预测的均方根误差 RMSE、正规化均方误差 NMSE 和相关系数 r 见表 5.1。同时为了说明数据长度对预测效果的影响，表 5.1 中多变量 1 表示用 $\{(x_{1,n},x_{2,n})\}_{n=1}^{2000}$ 预测 $\{x_{1,n}\}_{n=2001}^{2100}$，多变量 2 减少了一半数据，即用 $\{(x_{1,n},x_{2,n})\}_{n=1}^{2000}$ 中后 1000 个数据 $\{x_{1,n},x_{2,n}\}_{n=1001}^{2000}$ 预测 $\{x_{1,n}\}_{n=2001}^{2100}$。局部平均预测法和局部线性预测法中取最近邻点数 $K=12$，神经网络预测法中网络的隐层数为 8，训练次数为 1000 次[155]。

表 5.1　各种预测方法预测误差比较

预测方法		均方根误差 RMSE	正规化均方误差 NMSE	相关系数 r
局部平均预测法	单变量	0.2297	8.5133×10^{-4}	0.9896
	多变量 1	0.1260	2.5620×10^{-4}	0.9899
	多变量 2	0.1792	5.1807×10^{-4}	0.9898
局部线性预测法	单变量	0.0166	4.6931×10^{-6}	0.9900
	多变量 1	0.0027	1.1364×10^{-7}	0.9900
	多变量 2	0.0040	2.5874×10^{-7}	0.9900
神经网络预测法	单变量	0.0424	2.9041×10^{-5}	0.9900
	多变量 1	0.0037	2.2178×10^{-7}	0.9900
	多变量 2	0.0047	3.1490×10^{-7}	0.9900
径向基函数预测法	单变量	0.0021	9.5542×10^{-8}	0.9900
	多变量 1	0.0013	3.3002×10^{-8}	0.9900
	多变量 2	0.0018	6.3321×10^{-8}	0.9900

从表 5.1 可以得出：在已知数据长度相等的情况下，无论用什么预测方法，多变量重构预测都要比单变量重构预测好得多，即使多变量重构时只用单变量重构的一半数据，多变量预测也比单变量预测要好。因此，在历史数据长度比较短的情况下，利用多变量重构也有可能获得较高的预测精度。

为了更具体地说明选择不同的多变量对预测效果的影响，我们以用神经网络预测法预测变量 $x_{1,n}$ 的最后 100 个值为例详细说明。这里取隐层节点数是 8，迭代进行 1000 次。对单变量时间序列重构 $(x_{1,n},x_{1,n-4},x_{1,n-8})$、两变量时间序列重构 $(x_{1,n},x_{1,n-4},x_{2,n})$ 和 $(x_{1,n},x_{1,n-4},x_{3,n})$ 分别进行预测，预测的均方根误差 RMSE、正规化均方误差 NMSE 和相关系数 r 见表 5.2。

表 5.2　不同变量下神经网络预测法预测误差比较

变量	均方根误差 RMSE	正规化均方误差 NMSE	相关系数 r
$(x_{1,n},x_{1,n-4},x_{1,n-8})$	2.9041×10^{-5}	0.0424	0.9900
$(x_{1,n},x_{1,n-4},x_{2,n})$	2.2178×10^{-7}	0.0037	0.9900
$(x_{1,n},x_{1,n-4},x_{3,n})$	8.8101×10^{-6}	0.0234	0.9900

从表 5.2 中数据可见，无论是用两变量($x_{1,n}$，$x_{2,n}$)重构还是用两变量($x_{1,n}$，$x_{3,n}$)重构，预测的效果都好于仅用单变量 $x_{1,n}$ 预测的效果；用两变量($x_{1,n}$，$x_{2,n}$)重构的状态向量($x_{1,n}$，$x_{1,n-4}$，$x_{2,n}$)预测的正规化均方误差比仅用单变量 $x_{1,n}$ 重构的状态向量($x_{1,n}$，$x_{1,n-4}$，$x_{1,n-8}$)预测的正规化均方误差低两个数量级，这说明了适当选择多变量做预测可明显提高预测效果。另外，同样是用两变量，但用两变量($x_{1,n}$，$x_{2,n}$)重构要比用两变量($x_{1,n}$，$x_{3,n}$)重构预测好，所以适当选择变量对预测也是很重要的。

最后来比较多变量多项式预测与其他预测方法的预测效果。对单变量时间序列 $\{x_n\}_{n=1}^{2000}$，做式(5.63)的相空间重构。取 $s=2$，$p=1$，得单变量时间序列一步预测的二阶多项式预测模型

$$
\begin{aligned}
\hat{x}_{1,T+p} &= \hat{G}_1(\boldsymbol{x}_T) \\
&= c_0 + c_{10}^{(1)}x_{1,T} + c_{11}^{(1)}x_{1,T-4} + c_{12}^{(1)}x_{1,T-8} + c_{200}^{(1)}x_{1,T}^2 + c_{201}^{(1)}x_{1,T}x_{1,T-4} \\
&\quad + c_{202}^{(1)}x_{1,T}x_{1,T-8} + c_{211}^{(1)}x_{1,T-4}^2 + c_{212}^{(1)}x_{1,T-4}x_{1,T-8} + c_{222}^{(1)}x_{1,T-8}^2
\end{aligned} \tag{5.65}
$$

对两变量时间序列 $\{(x_n,y_n)\}_{n=1}^{2000}$，作式(5.64)的相空间的重构。取 $s=2$，$p=1$ 得两变量时间序列一步预测的二阶多项式预测模型

$$
\begin{aligned}
\hat{x}_{1,T+p} &= \hat{G}_1(\boldsymbol{x}_T) \\
&= c_0 + c_{10}^{(1)}x_{1,T} + c_{11}^{(1)}x_{1,T-4} + c_{200}^{(1)}x_{1,T}^2 + c_{201}^{(1)}x_{1,T}x_{1,T-4} + c_{211}^{(1)}x_{1,T-4}^2 \\
&\quad + c_{10}^{(2)}x_{2,T} + c_{200}^{(2)}x_{2,T}^2
\end{aligned} \tag{5.66}
$$

取最近邻点数 $K=100$，预测集的数据总数 $P=100$，预测误差如表 5.3 所示。

表 5.3　单变量与多变量时间序列局部多项式预测法预测结果比较

模型	非线性时间序列	均方根误差 RMSE	正规化均方误差 NMSE
局部多项式预测法	单变量 x_n	0.0071	8.0314×10^{-7}
	两变量(x_n，y_n)	0.0012	2.2941×10^{-8}

从表 5.3 的结果可知，基于多变量重构相空间的局部多项式法预测效果明显优于基于单变量重构相空间的局部多项式法的预测效果。进一步，将多变量时间序列局域预测法中三种预测方法对以上同组数据的预测结果进行比较，预测结果如表 5.4 所示。

表 5.4　多变量时间序列三种局域预测法预测结果比较

模型	均方根误差 RMSE	正规化均方误差 NMSE
局部平均预测法	0.5421	0.0047
局部线性预测法	0.1400	3.1360×10^{-6}
局部多项式预测法	0.0012	2.2941×10^{-8}

从表 5.4 的结果知，基于多变量重构相空间的局部多项式法预测效果明显优于局部平均预测法和局部线性预测法的预测效果。因此，给出的多变量时间序列的局部多项式预测法是一种更好的局部预测方法。

5.4 基于正则化的多变量非线性时间序列预测方法

在前面所研究的局部线性预测法、局部多项式预测法和径向基函数预测法中，都遇到了如下的线性回归问题：

$$y = Xc + e \tag{5.67}$$

式中，e 是误差向量，满足 $E(e)=0$，$\mathrm{Var}(e)=\sigma^2 I$。

当 X 的列向量线性无关时，用普通最小二乘法获得待求向量 c 的无偏估计，得到该问题的唯一解

$$\hat{c}_{\mathrm{OLS}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y \tag{5.68}$$

当 X 的列向量线性相关时，$X^{\mathrm{T}}X$ 是奇异的，此时 c 的估计不唯一。然而，这种情况在应用中很少见，更一般的情形是 X 的列向量近似线性相关，即存在多重共线性。正如 X 的列向量线性相关导致 $X^{\mathrm{T}}X$ 的奇异性，X 的列向量近似线性相关就会导致 $X^{\mathrm{T}}X$ 的近似奇异性。

奇异性可以定义为，存在一个单位向量 c（即 $\|c\|_2^2=1$），使得 $Xc=0$ 或写成 $c^{\mathrm{T}}X^{\mathrm{T}}Xc=0$。我们可以给近似奇异性归纳出如下特点，就是存在一个单位向量 c 使得 $\|Xc\|_2^2=c^{\mathrm{T}}X^{\mathrm{T}}Xc$ 非常小。当近似奇异性存在时，最小二乘估计的方差会非常大，估计的效果非常差。为了解决这一问题，下面给出改进的最小二乘估计，即正则化估计。

为了解释多重共线性导致最小二乘估计变坏的原因，也为正则化估计做准备，先给出矩阵的奇异值分解的概念。

5.4.1 奇异值分解

任何一个矩阵 X 都能写成三个简单矩阵的乘积，称为 X 的奇异值分解（singular value decomposition，SVD），这就是下面的奇异值分解定理[156,157]。

奇异值分解定理：若 $X\in R^{k\times m}$，则存在矩阵 $U=[u_1,u_2,\cdots,u_r]\in R^{k\times r}$ 和 $V=[v_1,v_2,\cdots,v_r]\in R^{m\times r}$，使

$$X = U\Sigma V^{\mathrm{T}} \tag{5.69}$$

其中 $U^{\mathrm{T}}U=I$，$V^{\mathrm{T}}V=I$，$\Sigma=\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$，且 $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_r>0$，$r=\min\{k,m\}$。

由奇异值分解定理可得

$$X=[u_1,u_2,\cdots,u_r]\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix}\begin{bmatrix}v_1^T\\ v_2^T\\ \vdots \\ v_r^T\end{bmatrix}=\sum_{i=1}^{r}\sigma_i u_i v_i^T \tag{5.70}$$

值空间

$$R(X)=\mathrm{span}\{u_1,u_2,\cdots,u_r\}\subset R^k \tag{5.71}$$

$$R(X^T)=\mathrm{span}\{v_1,v_2,\cdots,v_r\}\subset R^m \tag{5.72}$$

通常将子空间 $R(X)$ 看成输出空间；$R(X^T)$ 看成输入空间。

式(5.70)完全揭示了矩阵 X 作用于一个向量 c 所产生的多重共线性。例如，取 $c=v_j$，由于 $U^TU=I$，即

$$v_i^Tv_j=\begin{cases}0,当 i\neq j 时\\ 1,当 i=j 时\end{cases} \tag{5.73}$$

从而有

$$Xv_j=\sum_{i=1}^{r}\sigma_i u_i v_i^T v_j=\sigma_j u_j \tag{5.74}$$

更一般地，设 $c=\alpha_1v_1+\alpha_2v_2+\cdots+\alpha_rv_r$，则

$$Xc=\sum_{i=1}^{r}\sigma_i u_i v_i^T\left(\sum_{j=1}^{r}\alpha_j v_j\right)=\sum_{i=1}^{r}\sigma_i\alpha_i u_i \tag{5.75}$$

因此，向量 c 在方向 v_j 上的一个微小变化仅对乘积 Xc 在方向 u_j 上产生影响。可以说成输入空间中的每个方向 v_j 联系着输出空间中的一个特定方向 u_j，这种联系的紧密程度完全取决于相应的奇异值 σ_j。例如，若 $\sigma_j=1000$，则输入方向 v_j 上的一个微小变化能引起输出方向 u_j 上的一个巨大变化。若 $\sigma_j=0$，则输入方向 v_j 上的变化对输出方向 u_j 没有任何影响。

显然，最大与最小奇异值之比总是大于等于 1，称这个比值为矩阵 X 的条件数，记为 γ，即

$$\gamma=\frac{\sigma_1}{\sigma_r} \tag{5.76}$$

一个非常大的条件数隐含着某些奇异值非常小(与 σ_1 相比较)。假设最小的奇异值满足 $\sigma_r\approx 0$，取 $c=v_r$，也就是“输入空间”中的方向 v_r 对应奇异值 σ_r，那么由 $Xv_r=\sigma_r u_r$，得到

$$\|Xc\|_2=\|Xv_r\|_2=\|\sigma_r u_r\|_2=\sigma_r\|u_r\|_2=\sigma_r\approx 0 \tag{5.77}$$

这就精确地表达了前面提到的多重共线性。换句话说，多重共线性意味着 X 有一个非常小的奇异值，或者是 X 的条件数非常大。

5.4.2 最小二乘估计

下面将在奇异值分解的框架内推导出普通最小二乘(ordinary least square,

OLS)估计，找出 X^TX 近似奇异时导致最小二乘估计变坏的原因，以便对它进行改进。

1. 最小二乘估计的奇异值分解表示

对线性回归模型 $y=Xc+e$，将矩阵 X 的奇异值分解 $X=U\Sigma V^T$ 代入它的最小二乘估计 $\hat{c}_{OLS}=(X^TX)^{-1}X^Ty$ 中，得到

$$\hat{c}_{OLS}=V\Sigma^{-1}U^Ty \tag{5.78}$$

也可写成

$$\hat{c}_{OLS}=\sum_{i=1}^{r}\frac{(u_i,y)}{\sigma_i}v_i \tag{5.79}$$

其中$(u_i,y)=u_i^Ty\ (i=1,2,\cdots,r)$是 y 投影到单位向量 u_i 上的长度，叫做 y 的傅里叶系数。

根据 Gauss-Markov 定理，最小二乘估计在线性无偏估计类中方差最小。因此，为了说明什么情况下线性无偏估计的性质明显变坏，提出一个度量估计优劣的另一个标准——均方误差。

2. 最小二乘估计的均方误差

定义均方误差

$$\mathrm{MSE}(\hat{c})=E[\|\hat{c}-c\|_2^2] \tag{5.80}$$

这里采用的是关于 $\hat{c}$ 的随机误差的期望，式(5.80)可改写成

$$\begin{aligned}\mathrm{MSE}(\hat{c})&=E[\|\hat{c}-E[\hat{c}]+E[\hat{c}]-c\|_2^2]\\&=E[\|\hat{c}-E[\hat{c}]\|_2^2]+2E[(\hat{c}-E[\hat{c}])^T(E[\hat{c}]-c)]+E[\|E[\hat{c}]-c\|_2^2]\\&=E[\|\hat{c}-E[\hat{c}]\|_2^2]+\|E[\hat{c}]-c\|_2^2\end{aligned} \tag{5.81}$$

记

$$\mathrm{bias}(\hat{c})=E[\hat{c}]-c \tag{5.82}$$

$$\mathrm{Var}(\hat{c})=E[\|\hat{c}-E[\hat{c}]\|_2^2] \tag{5.83}$$

则有

$$\mathrm{MSE}(\hat{c})=\mathrm{Var}(\hat{c})+\|\mathrm{bias}(\hat{c})\|_2^2 \tag{5.84}$$

由式(5.82)，$\mathrm{bias}(\hat{c}_{OLS})=0$，由式(5.83)

$$\begin{aligned}\mathrm{Var}(\hat{c}_{OLS})&=\mathrm{trCov}(\hat{c}_{OLS})=\mathrm{trCov}(V\Sigma^{-1}Uy)\\&=\mathrm{tr}[V\Sigma^{-1}U^T\mathrm{Cov}(y)U\Sigma^{-1}V^T]\\&=\mathrm{tr}[\sigma^2V\Sigma^{-2}V^T]\\&=\sigma^2\sum_{i=1}^{r}\frac{1}{\sigma_i^2}\end{aligned} \tag{5.85}$$

式中，r 是矩阵 X 的秩；σ^2 是基本模型 $y=Xc+e$ 中噪声项 e 的方差，计算中用到了

$\mathrm{Cov}(\boldsymbol{y})=\sigma^2 I$。于是最小二乘估计的均方误差为

$$\mathrm{MSE}(\hat{c}_{\mathrm{OLS}})=\sigma^2\sum_{i=1}^{r}\frac{1}{\sigma_i^2} \tag{5.86}$$

由式(5.86)，当 X 的列向量间存在多重共线性时，X 至少有一个奇异值 σ_i 非常小，这就导致 $\hat{c}_{\mathrm{OLS}}$的均方误差非常大，由(5.84)知也即等价于 $\hat{c}_{\mathrm{OLS}}$的方差非常大，从而 $\hat{c}_{\mathrm{OLS}}$估计变坏。于是，人们开始考虑最小二乘估计的改进，也即寻找新的估计来减小方差，而根据 Gauss-Markov 定理，最小二乘估计在线性无偏估计类中方差最小，因此在这种意义下没有其他的无偏估计方差会更小。Gauss-Markov 定理并没有提及有偏估计和非线性估计。所以，下面将考虑几种有偏估计，也称正则化估计，它能达到减少均方误差的目的。

5.4.3 正则化估计

由式(5.84)，均方误差是方差与偏差平方之和，而且最小二乘估计的偏差为0，因此要想获得比最小二乘估计更小的均方误差，必须在减少方差的同时避免引入同等数量的偏差。由式(5.79)，即 $\hat{c}_{\mathrm{OLS}}=\sum_{i=1}^{r}\frac{(\boldsymbol{u}_i,\boldsymbol{y})}{\sigma_i}\boldsymbol{v}_i$ 可以看出，较小的奇异值 σ_i 对最小二乘估计的影响较大，为了保持估计的稳健性，应该降低这样一些方向的权值。为此，对最小二乘估计作如下修改

$$\hat{\boldsymbol{c}}=\sum_{i=1}^{r}\lambda_i\frac{(\boldsymbol{u}_i,\boldsymbol{y})}{\sigma_i}\boldsymbol{v}_i \tag{5.87}$$

它与最小二乘估计的唯一不同之处是在每个方向引入了一个权值 λ_i，称作过滤因子。若 σ_i 较大，对这样一些方向我们让 $\lambda_i=1$；若 σ_i 较小，对这样一些方向就选择较小的 λ_i，小的过滤因子会消除小奇异值的影响，从而均方误差相对于最小二乘估计而言就会减少。下面给出估计式(5.87)的均方误差表达式。

将式(5.87)写成矩阵形式：

$$\hat{\boldsymbol{c}}=\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{y} \tag{5.88}$$

其中

$$\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_r)$$

由于$E[\hat{\boldsymbol{c}}]=\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}}E[\boldsymbol{y}]=\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{c}$，而 $\boldsymbol{c}$ 的傅里叶展开式为 $\hat{\boldsymbol{c}}=\sum_{i=1}^{r}(\boldsymbol{v}_i,\boldsymbol{c})\boldsymbol{v}_i$，这里$(\boldsymbol{v}_i,\boldsymbol{c})$是 $\boldsymbol{c}$ 的傅里叶系数，所以

$$\begin{aligned}\|\mathrm{bias}(\hat{\boldsymbol{c}})\|_2^2&=\|E[\hat{\boldsymbol{c}}]-\boldsymbol{c}\|_2^2=\|(\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}-\boldsymbol{I})\boldsymbol{c}\|_2^2\\&=\sum_{i=1}^{r}(\lambda_i-1)^2(\boldsymbol{v}_i,\boldsymbol{c})^2\end{aligned} \tag{5.89}$$

又

$$\begin{aligned}\mathrm{Var}(\hat{c}) &= \mathrm{trCov}(\hat{c}) = \mathrm{tr}[\boldsymbol{V\Sigma}^{-1}\boldsymbol{\Lambda U}^{\mathrm{T}}\mathrm{Cov}(\boldsymbol{y})\boldsymbol{U\Lambda\Sigma}^{-1}\boldsymbol{V}^{\mathrm{T}}] \\ &= \mathrm{tr}[\sigma^2\boldsymbol{V\Sigma}^{-2}\boldsymbol{\Lambda}^2\boldsymbol{V}^{\mathrm{T}}] \\ &= \sigma^2\sum_{i=1}^{r}\frac{\lambda_i^2}{\sigma_i^2}\end{aligned} \tag{5.90}$$

于是估计式(5.87)的均方误差为

$$\mathrm{MSE}(\hat{c}) = \sum_{i=1}^{r}(\lambda_i - 1)^2(\boldsymbol{v}_i,\boldsymbol{c})^2 + \sigma^2\sum_{i=1}^{r}\frac{\lambda_i^2}{\sigma_i^2} \tag{5.91}$$

对最小二乘估计，过滤因子 $\lambda_i=1, i=1,\cdots,r$，所以式(5.91)中的第一项为零，从而 $\mathrm{MSE}(\hat{c})$减少到式(5.86)。从式(5.91)中还可以看出，当过滤因子 λ_i 从 1 开始增加时，偏差项和方差项都随之增加，因此 $\mathrm{MSE}(\hat{c})$变大，故应控制 λ_i 的取值，使 $0\leqslant\lambda_i\leqslant1$；当 λ_i 从 1 减小到 0 时，偏差项增大了，但方差项变小了，所以在 0～1 中寻找最佳的 λ_i 值，使 $\mathrm{MSE}(\hat{c})$达到最小是有待研究的问题。

下面结合 λ_i 的取值介绍几种正则化估计，它们都具有比最小二乘估计更小的均方误差。

1. 主成分回归估计

由于最小的一些奇异值使得最小二乘估计方差变大，为了降低这种影响，从最小二乘估计的表达式中去掉这些项，也就是在式(5.79)的和式中去掉第 q 项之后的项(因为最后的这些项对应最小的奇异值)，对应的估计叫做主成分回归(principal components regression，PCR)估计[150]，记作

$$\hat{c}_{\mathrm{PCR}} = \sum_{i=1}^{q}\frac{(\boldsymbol{u}_i,\boldsymbol{y})}{\sigma_i}\boldsymbol{v}_i \tag{5.92}$$

注意到对应截断点 q 的不同选择，我们能得到许多 PCR 估计。对应截断点为 q 的 PCR 估计记成 PCR(q)，数 q 在估计中常称作因子数。

易见 PCR(q)估计的过滤因子为

$$\lambda_i = \begin{cases}1, 1\leqslant i\leqslant q \\ 0, q+1\leqslant i\leqslant r\end{cases} \tag{5.93}$$

将其代入式(5.91)，得到 PCR(q)估计的均方误差为

$$\mathrm{MSE}(\hat{c}_{\mathrm{PCR}(q)}) = \sum_{i=1}^{r}(\boldsymbol{v}_i,\boldsymbol{c})^2 + \sigma^2\sum_{i=1}^{q}\frac{1}{\sigma_i^2} \tag{5.94}$$

要使

$$\mathrm{MSE}(\hat{c}_{\mathrm{PCR}(q)}) < \mathrm{MSE}(\hat{c}_{\mathrm{OLS}}) \tag{5.95}$$

只须

$$\sum_{i=q+1}^{r}(\boldsymbol{v}_i,\boldsymbol{c})^2 < \sigma^2\sum_{i=q+1}^{r}\frac{1}{\sigma_i^2} \tag{5.96}$$

对于 PCR(q)估计，由于 X 的后 $r-q$ 个奇异值接近于零，所以式(5.96)右端非常大，即不等式(5.96)总是成立的。也就是说，PCR(q)估计的均方误差总比最小二乘估计的均方误差小。

对于 PCR(q)估计，参数 q 的选取是关键。其中最常用的是交叉校验(cross validation, CV)和广义交叉校验(GCV)[159]；还有更简单确定 q 的方法，就是利用奇异值谱(奇异值谱中明显的间隙处对应的值取作 q)，或要求 X 的某些奇异值之和占整个奇异值之和的比例达到预先给定的值(一般取大于等于 80%)[160]。在混沌时间序列的应用中，CV 常常会过估正则化参数。至于利用奇异值谱选取 PCR 的参数 q，在谱中很少会出现一个明显的间隙。利用比例计算 q 也不是一个很严格的标准，这是由于 q 的估计会随着 m 一起改变。文献[157]给出了一个较好的方法，即对非线性动力系统，如果系统中吸引子的拓扑维 d 能够估计出来，则对固定的 q 值建议取

$$q=[d] \tag{5.97}$$

式中：$[d]$是大于 d 的最小整数。

2. 岭回归估计

最小二乘估计是对问题 $\|y-Xc\|_2^2$ 极小化得到的，在多重共线性情况下，式(5.79)中某些 σ_i 非常小，所以 $\hat{c}_{\text{OLS}}$的范数非常大，为避免这种情况发生，我们改成极小化问题

$$H(c)\|y-Xc\|_2^2+\mu\|c\|_2^2 \tag{5.98}$$

式中，μ 是一个正常数。这样做的目的是惩罚大范数 $\|c\|$，迫使解的范数变小。由此得到的估计叫做岭回归(ridge regression, RR)估计[161]。容易得到 RR 估计为

$$\hat{c}_{\text{RR}}=(X^{\text{T}}X+\mu I)^{-1}X^{\text{T}}y \tag{5.99}$$

式中，I 为单位阵。将 X 的奇异值分解 $X=U\Sigma V^{\text{T}}$ 代入式(5.99)，得

$$\begin{aligned}\hat{c}_{\text{RR}}&=(V\Sigma^2V^{\text{T}}+\mu I)^{-1}V\Sigma U^{\text{T}}y\\&=\sum_{i=1}^{r}\frac{\sigma_i}{\sigma_i^2+\mu}(u_i,y)v_i\end{aligned} \tag{5.100}$$

RR 估计的过滤因子为

$$\lambda_i=\frac{\sigma_i^2}{\sigma_i^2+\mu} \tag{5.101}$$

因为 $\mu>0$，故 $0\leqslant\lambda_i<1$，从而 RR 估计在每个方向上都对最小二乘估计进行了压缩。将式(5.100)代入式(5.91)，得到 RR 估计的均方误差为

$$\text{MSE}(\hat{c}_{RR})=\sum_{i=1}^{r}\frac{\mu^2}{(\sigma_i^2+\mu)^2}(v_i,b)^2+\sigma^2\sum_{i=1}^{r}\frac{\mu^2}{(\sigma_i^2+\mu)^2} \tag{5.102}$$

对找到的一个 μ，当 c 和 σ^2 满足一定条件时有，$\text{MSE}(\hat{c}_{\text{RR}})<\text{MSE}(\hat{c}_{\text{OLS}})$成立。

对于 RR 估计，参数 μ 的选取是关键。对 μ 可简单取作残差方差的估计 $\hat{\sigma}^2$，残差方差的估计为

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{k-r}\hat{e}^{\mathrm{T}}\hat{e} = \frac{1}{k-r}\| y - X\hat{c} \|_2^2 = \frac{1}{k-r}\| U^{\mathrm{T}}y - U^{\mathrm{T}}X\hat{c} \|_2^2 \\
&= \frac{1}{k-r}\| U^{\mathrm{T}}y - U^{\mathrm{T}}XVV^{\mathrm{T}}\hat{c} \|_2^2 \\
&= \frac{1}{k-r}\| U^{\mathrm{T}}y - \Sigma V^{\mathrm{T}}\hat{c} \|_2^2 \\
&= \frac{1}{k-r}\Big[\sum_{i=1}^{r}(u_i^{\mathrm{T}}y - \sigma_i v_i^{\mathrm{T}}\hat{c}) + \sum_{i=r+1}^{k}(u_i^{\mathrm{T}}y)^2 \Big] \\
&= \frac{1}{k-r}\sum_{i=r+1}^{k}(u_i^{\mathrm{T}}y)^2
\end{aligned}
\tag{5.103}
$$

3．其他正则化估计

正则化估计还有偏最小二乘估计[161]、特征根估计等。如果能适当引入某种过滤因子使得相应的均方误差比最小二乘估计的均方误差 MSE($\hat{c}_{\mathrm{OLS}}$)小，还可以得到新的正则化估计。

5.4.4　基于正则化的局部线性和局部多项式预测的步骤

基于前面的研究，下面将从消除多重共线性的角度引入正则化参数，对多变量局部线性预测法和局部多项式预测法进行改进，最终提高非线性时间预测法的预测精度。

多变量局部线性预测法和局部多项式预测法，均是先建立模型，然后用最小二乘估计进行模型参数的估计。在这两种局部预测法模型建立中都需要选取邻点 K，目前还没有统一的选取方法，对局部线性预测法常用的方法是文献[123]中提出的方法，这种方法实际上确定了 K 在 $2m+1$ 至 $2m+10$ 中取得，也即 $K \gg m$。对局部多项式预测法为了得到稳定的解，一般要求满足 $K \geqslant J$，其中 J 为由式(5.24)给出的待定系数的个数，与多项式的阶数 s 有关。实际上，将非线性时间序列映射到高维相空间中，尽管有利于消除噪声等因素造成的“伪邻近点”，但由于实际问题中样本总是有限的，因此有时在相空间中某些点上却不能找到足够数目的邻近点，因而有时条件 $K \gg m$(局部线性预测时)或 $K \geqslant J$(局部多项式预测时)并不一定能够达到。当 K 接近 m 或 J 时，最小二乘估计变得很不稳定，这时正则化估计的作用很明显。尤其当 $K \leqslant m$ 或 $K \leqslant J$ 时，正则化估计中的 PCR 估计的作用尤其明显，这是因为即使相空间的维数 m 较大，回归问题的实际维数仍较稳定，大小与 q 相同，因此由 PCR 获得的解是稳定的。

另外，由于实际问题所观测到的数据中总是含有噪声的，当观测数据中引入了

测量噪声时，尽管这时矩阵 X 的条件数总是较适中的，但由于 OLS 解中的部分方向被噪声覆盖，未来的状态没有包含真正有用的信息，所以 OLS 的预测能力退化了，而正则化估计可以有效地过滤掉解中的这些方向。因此，正则化估计对于实际问题具有明显的改进作用。

下面将对多变量局部线性预测法和局部多项式预测法进行改进，得到新的预测方法。其步骤如下：

第一步：根据 4.3 节给出的方法选取时间延迟 τ，嵌入维数 m，重构相空间。

第二步：给定一个合适的邻近点数目 K，并找出与当前点最近的 K 个邻近点。

第三步：根据式(5.97)及式(5.103)确定正则化参数 q 或 μ 的值。

第四步：根据式(5.79)、式(5.92)及式(5.100)分别做出 c 的估计。

第五步：对多变量局部线性预测，根据式(5.8)式做出预测；对多变量局部多项式预测，根据式(5.19)做出预测。

第六步：对新的目标点，重复第一步至第五步直到结束。

以上基于正则化回归的一般局部线性预测和局部多项式预测与基于最小二乘回归的一般局部线性预测和局部多项式预测的方法是相同的，区别仅在于对回归问题的参数选用的估计式不同，对于不同的研究问题，它们各具优势。

考虑一列来自于低维系统的不含噪声的数据，当它嵌入的相空间维数 m 足够大时，吸引子会在某些方向上具有局部小的方差，此时数据矩阵 X 是病态的。例如，对 Henon 映射产生的时间序列($K=15, m=5$)，X 的条件数 σ_1/σ_5 可以达到 $O(10^4)$。这是由于系统的吸引子主要集中在 R^m 的子空间中，而该子空间外部的微小变化可能包含较为有用的信息，然而正则化估计会将这样一些信息过滤掉，这就导致正则化所作的预测变坏。只要数据矩阵的条件数不大到会产生数值计算方面的问题时，普通最小二乘估计仍然是最好的。另外，当 $K \leqslant m$ 时，OLS 估计变得不稳定，这时正则化的作用非常明显。

5.4.5 Lorenz 系统的仿真模拟

为了验证多变量非线性时间序列的正则化局部线性和局部多项式预测法优于一般的局部线性和局部多项式预测法，下面采用分别加入 5%和 10%正规白噪声的 Lorenz 系统进行仿真检验。考虑以下 Lorenz 系统

$$\left.\begin{aligned}\frac{\mathrm{d}x_1}{\mathrm{d}t} &= \sigma(x_2 - x_1)\\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= x_1(r - x_3) - x_2\\ \frac{\mathrm{d}x_3}{\mathrm{d}t} &= x_1x_2 - bx_3\end{aligned}\right\} \tag{5.104}$$

其中 $\sigma=10, r=28, b=\frac{8}{3}$。取 $x_{1,0}=15.34, x_{2,0}=13.68, x_{3,0}=37.91$，积分步长 $h=0.04$。利用四阶 Runge-Kutta 积分法计算出两个包含 10 000 个数据的仿真时间序列 x_1 和 x_2。取 5001～6000 之间的 $N=1000$ 个数据参加实验，得到多变量时间序列 $\{(x_{1,n}, x_{2,n})\}_{n=1}^{1000}$。由 4.3.4 节给出的算法得 $\tau_1=\tau_2=4, m_1=2, m_2=2$。对主成分回归(PCR)估计，由于 Lorenz 吸引子的拓扑维 $d=2.06$，故取正则化参数 $q=3$；对岭回归(RR)估计，取 $\mu=\hat{\sigma}^2$。

1. *基于正则化的局部线性预测*

图 5.4 表示对 Lorenz 系统的两变量非线性时间序列 $\{(x_{1,n}, x_{2,n})\}_{n=1}^{1000}$ 加 5%噪声后，利用普通最小二乘的局部线性预测法(OLS)和利用正则化的局部线性预测法(PCR, RR)，对最后 100 个数据进行一步预测得到的正规化均方根误差 NRMSE。

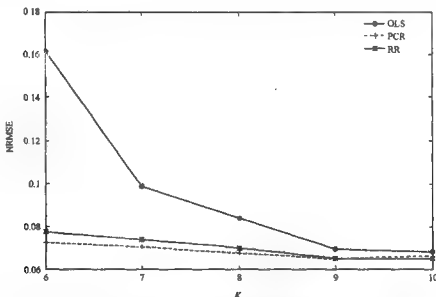


图 5.4　加 5%噪声后 Lorenz 系统的局部线性预测的一步预测误差

图 5.5 表示对 Lorenz 系统的两变量非线性时间序列 $\{(x_{1,n}, x_{2,n})\}_{n=1}^{1000}$ 加 10%噪声后，利用普通最小二乘的局部线性预测法(OLS)和利用正则化的局部线性预测法(PCR, RR)，对最后 100 个数据进行一步预测得到的正规化均方根误差 NRMSE。

表 5.5 给出了 Lorenz 系统的两变量非线性时间序列 $\{(x_{1,n}, x_{2,n})\}_{n=1}^{1000}$，利用普通最小二乘的局部线性预测法(OLS)和正则化的局部线性预测法(PCR, RR)在

不同噪声级别及最近邻点数 K 下对最后 100 个数据进行一步预测的正规化均方根误差 NRMSE 的比较。

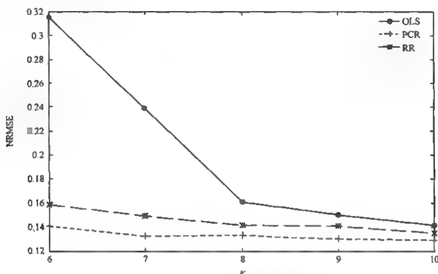


图 5.5　加 10% 噪声后 Lorenz 系统的局部线性预测的一步预测误差

表 5.5　含噪声的 Lorenz 系统的局部线性预测的正规化均方根误差 NRMSE

噪声级别	预测方法	$K>m+1$						$K=m+1$	$K\leqslant m$
		10	9	8	7	6	5	4	3
5%	OLS	0.0680	0.0693	0.0852	0.0983	0.1613	1.4898	—	—
	PCR	0.0662	0.0649	0.0676	0.0705	0.0723	0.0844	0.1471	0.1132
	RR	0.0646	0.0647	0.0699	0.0734	0.0769	—	—	—
10%	OLS	0.1415	0.1503	0.1610	0.2397	0.3142	1.9085	—	—
	PCR	0.1298	0.1304	0.1333	0.1326	0.1407	0.1494	0.2901	0.2671
	RR	0.1347	0.1405	0.1414	0.1488	0.1587	—	—	—

由图 5.4、图 5.5 和表 5.5 可以看出，对含不同噪声级别的多变量非线性时间序列进行局部线性一步预测时，都有同样的结果，即

1) 当 $K>m+1$ 时，所有的正则化预测法都优于普通最小二乘预测法。

2) 使用基于普通最小二乘法的局部线性化方法时，K 的取值对结果影响非常大，当 K 取较大的值时，精确度提高了，但随着 m 接近 K，误差增长较快，且一般较小的 K 值所计算出的误差较大，特别当 K 与 m 非常接近时(见 $K=m+1$ 情形)，误差急剧上升。当 $K\leqslant m$ 时，预测失败。

3) 当 K 与 m 非常接近时，甚至 $K\leqslant m$ 时，正则化的 PCR 法仍有较好的预测效果。

4）当增加噪声的级别后，基于普通最小二乘和基于正则化的局部线性预测法的预测误差都有所增大，但是基于普通最小二乘的局部线性预测法的预测误差增加的幅度最大。

2. 基于正则化的局部多项式预测

利用式(5.19)进行局部多项式预测，此时取多项式阶数 $s=2$，则待定系数个数 $J=\sum_{l=1}^{2}C_{m_l+2}^{2}-2+1=C_4^2+C_4^2-2+1=11$。图 5.6 表示对 Lorenz 系统的两变量非线性时间序列 $\{(x_{1,n},x_{2,n})\}_{n=1}^{1000}$ 加 5%噪声后，利用普通最小二乘的局部多项式预测法(OLS)和利用正则化的局部多项式预测法(PCR，RR)，对最后 100 个数据进行一步预测得到的正规化均方根误差 NRMSE。

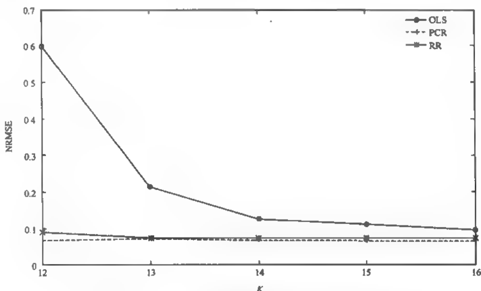


图 5.6 加 5%噪声后 Lorenz 系统的局部多项式预测的一步预测误差

图 5.7 表示对 Lorenz 系统的两变量非线性时间序列 $\{(x_{1,n},x_{2,n})\}_{n=1}^{1000}$ 加 10%噪声后，利用普通最小二乘的局部多项式预测法(OLS)和利用正则化的局部多项式预测法(PCR，RR)，对最后 100 个数据进行一步预测得到的正规化均方根误差 NRMSE。

表 5.6 给出了 Lorenz 系统的两变量非线性时间序列 $\{(x_{1,n},x_{2,n})\}_{n=1}^{1000}$，利用普通最小二乘的局部多项式预测法(OLS)和正则化的局部多项式预测法(PCR，RR)在不同噪声级别及最近邻点数 K 下对最后 100 个数据进行一步预测的正规化均方根误差 NRMSE 的比较。

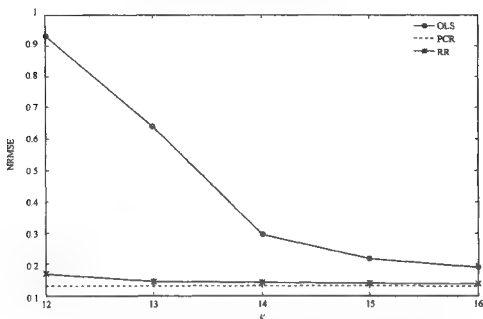


图 5.7　加 10%噪声后 Lorenz 系统的局部多项式预测的一步预测误差

表 5.6　含噪声的 Lorenz 系统的局部多项式预测的正规化均方根误差 NRMSE

噪声级别	预测方法	K≥J						K<J	
		16	15	14	13	12	11	10	9
5%	OLS	0.0946	0.1116	0.1257	0.2126	0.5960	7.403	1.1782	2.4834
	PCR	0.0644	0.0648	0.0655	0.0705	0.0655	0.0666	0.0673	0.0670
	RR	0.0704	0.0712	0.0731	0.0734	0.0864	—	—	—
10%	OLS	0.1890	0.2162	0.2940	0.6395	0.9273	3.4985	4.9619	6.3496
	PCR	0.1250	0.1277	0.1279	0.1288	0.1293	0.1324	0.1353	0.1388
	RR	0.1349	0.1377	0.1399	0.1446	0.1682	—	—	—

由图 5.6、图 5.7 和表 5.6 可以看出，对含不同噪声的级别的多变量非线性时间序列进行局部多项式一步预测时，与局部线性一步预测有类似的结果，即

1）当 $K>J+1$ 时，所有的正则化预测法都优于普通最小二乘预测法。

2）使用基于普通最小二乘法的局部多项式预测方法时，K 的取值对结果影响非常大，当 K 取较大的值时，精确度提高了，但随着 K 接近于待定系数个数 J 时，误差增长较快，且一般较小的 K 值所计算出的误差较大，特别当 K 与 J 非常接近时（见 $K=J$ 情形），误差急剧上升，预测误差大于 1，因而此时预测效果很差。

3）当 K 与 J 非常接近时，甚至 $K \leqslant J$ 时，正则化的 PCR 法仍有较好的预测效果。

4）当增加噪声的级别后，基于普通最小二乘法和基于正则化法的局部多项式预测方法的预测误差都有所增大，但是基于普通最小二乘法的局部多项式预测方法的预测误差增加的幅度最大。

由上面的分析表明，当 K 取得较大时，不管是基于普通最小二乘法还是基于正则化方法所得的预测误差都较小，当 K 取得较小时，基于正则化方法的预测误差要远小于基于普通最小二乘法的预测误差。实际问题中由于样本容量有限，K 值不可能无限增大，随着 K 值的增加计算量加大，所以基于正则化的多变量局部线性化法或局部多项式法预测效果比一般的局部线性化或局部多项式预测法好；基于正则化的局部线性预测法或局部多项式预测法抗噪性比一般的局部线性预测法或局部多项式预测法强；基于正则化的局部线性预测方法或局部多项式预测方法的预测范围比一般的局部线性化或局部多项式预测法广。因此，在实际问题中用基于正则化的局部线性化预测方法或局部多项式预测方法更有应用价值。

5.5　基于正则化的多变量非线性时间序列的自适应预测方法

一般的局部线性化预测方法是取固定的嵌入维数 m，邻近点数目 K 及正则化参数 q 与 μ。按照 F. Taken 的嵌入定理，宜选择较大的嵌入维数。将时间序列映射到高维相空间中，尽管有利于消除噪声等因素造成的虚假邻点，但是鉴于样本数量有限，在相空间中某些点处可能找不到足够数目的邻近点，这是导致预测误差增加的主要原因[92]；将混沌时间序列映射到较低维相空间中，尽管可能存在轨道交叉和虚假邻点，但是在相空间中容易找到当前点的足够数目的邻近点。因此，对不同的预测点，应选择不同的嵌入维数，以达到改善预测精度的目的。文献[163]给出了一种基于奇异值分解的自适应局部线性化方法，它可以自适应确定当前嵌入维数，从而克服病态数据矩阵的影响。该方法与一般的局部线性化方法的最大不同之处是，前者在每个预测时刻可采用不同的嵌入维数，而后者在所有的预测时刻均采用相同的嵌入维数。然而，文献[163]仅对最小二乘法进行了研究，且对邻近点数目 K 取的是固定值，这里在文献[163]的基础上对邻近点数目 K 及正则化参数 q 与 μ 均采用自适应取值，这样可以大大提高基于一般正则化回归所作的局部线性预测的精度。因为在当前时刻自适应地选择合适的嵌入维数 m、邻近点数目 K 及正则化参数 q 与 μ，可望获得比这些参数取固定值时更好的预测结果。

5.5.1　基于正则化的自适应预测的步骤

当 X 不是列满秩或 X 的条件数过大时，最小二乘法估计出的参数不可信，因

此需要改良 X，以使其列满秩且条件数不大于一个给定的正数 M，从而保证参数估计的稳健性。在当前时刻点，如果足够大的嵌入维数是合理的，那么数据矩阵 X 列满秩且其条件数不大于 M；反之，X 很可能是病态的。基于这个分析，我们可以在不损失估计精度的前提下达到此目的，做法是先选一个初始嵌入维数 m，然后在当前时刻点，如果 X 是病态的，就做降维处理，从而找到一个最大的使 X 列满秩且其条件数不大于 M 的嵌入维数 m_0[163]。算法步骤如下：

第一步：初选嵌入维数 m，并给定条件数阈值 $M>0$。

第二步：在当前时刻，由 5.1.2 节给出的方法自适应地确定邻近点数目 K，构造 X，并对 X 做 SVD 分解。

第三步：若 $\sigma_1/\sigma_r<M$，则取 $m_0=r$，否则转第四步。

第四步：$r=r-1$；$\sigma_r=\sigma_{r-1}$。

第五步：重复第三步和第四步直至 $\sigma_1/\sigma_r<M$。

确定了合适的嵌入维数 m_0 后，就可以重构相空间，然后用正则化回归方法计算在当前时刻点的预测值。其步骤如下：

第一步：初选嵌入维数 m，并给定条件数阈值 $M>0$。

第二步：由 5.1.2 节给出的方法自适应地确定邻近点数目 K。

第三步：在当前时刻，调用上面的算法自适应地确定合适的嵌入维数 m_0。

第四步：自适应地确定正则化参数 q 与 μ。

第五步：如果 $m_0=m$，则根据式(5.79)、式(5.92)及式(5.100)分别确定 c，然后根据式(5.8)做出预测。

第六步：否则，根据 m_0 重构相空间，再自适应地确定邻近点数目 k，并用式(5.79)及式(5.92)、式(5.100)做出预测。

第七步：对新的目标点，重复第一步至第六步直到结束。

上述算法中的条件数阈值 M 一般根据实验结果合理选择。根据经验，对许多实际问题，一般取 $M=100$ 较合适。

为了对自适应局部线性化法有一个初步认识，首先将由两种方法计算出的奇异值和过滤因子进行对比，来说明自适应局部线性化法的优势。以 Lorenz 系统为例，图 5.8～图 5.11 显示的是当 $m=8,\tau=4,k=15$ 时用一般局部线性化方法与自适应局部线性化方法分别对单个值进行一步预测时奇异值、傅里叶系数、残差标准差和过滤因子的结果。对主成分回归 PCR，取 $q=3$，对岭回归 RR，取 $\mu=\hat{\sigma}^2$。

由图 5.8 可以看出，残差标准差 $\hat{\sigma}$ 接近于零，并且最后四个 σ_i 及 $|u_i^{\mathrm{T}}y|$ 接近于零，由此可知最小二乘估计变坏，而对正则化估计而言，对应的过滤因子必须为零，才能消除小奇异值的影响。在图 5.9 中，PCR 的后面五个过滤因子全为零，所以取 $q=3$ 显得过于保守，过滤掉了方向 u_4，减少了可以利用的信息，进而会影响估计的准确性。对于 RR，由于残差标准差太小，致使后四个过滤因子全不为零，RR 估计

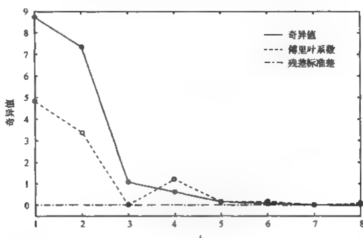


图 5.8　一般局部线性化方法计算的奇异值、傅里叶系数及残差标准差

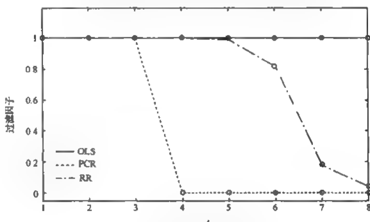


图 5.9　一般局部线性化方法计算的过滤因子

没有对 OLS 估计进行有效压缩。因此，由一般局部线性化法得到的过滤因子没有起到应有的作用。在相同条件下，当改用自适应局部线性化法时，结果有明显改进。通过自适应选取嵌入维 m，得到该点较合适的嵌入维数是 6，图 5.10 中，由自适应局部线性化法计算的残差标准差 $\hat{\sigma}$ 小于主要的奇异值 σ_i 及傅里叶系数的大 $|u_i^T y|$，并且最后两个 σ_i 及 $|u_i^T y|$ 接近于零，图 5.11 中正则化估计对应的过滤因子全为零。通过这个例子检验了自适应确定正则化参数的方法是合理的，因此可以预期由自适应局部线性化方法能够得到更好的预测结果。

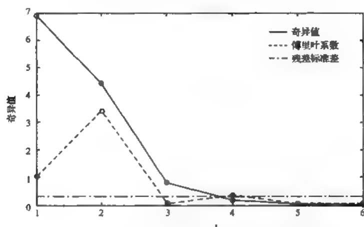


图 5.10　自适应局部线性化方法计算的奇异值、傅里叶系数及残差标准差

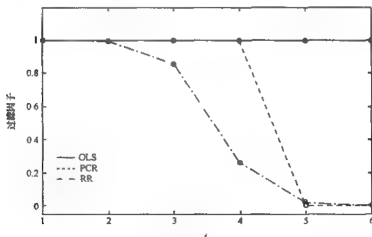


图 5.11　自适应局部线性化方法计算的过滤因子

5.5.2　Henon 映射的仿真检验

为了分析使用自适应局部线性化方法以后最小二乘回归和正则化回归的改进程度，下面将最小二乘回归和正则化回归放在一起进行对比研究。Henon 映射是一个具有吸引子的二维迭代映射，其方程如下：

$$\begin{cases} x_{n+1} = 1 + y_n - ax_n^2 \\ y_{n+1} = bx_n \end{cases} \tag{5.105}$$

当 $a=1.40$ 和 $b=0.30$ 时，获得了混沌运动。取初值(0.1, 0.0)进行迭代，得$\{x_n\}$的 1100 个采样点，并给它附加 5%的正规白噪声。分别用一般局部线性化方法以及自适应局部线性化方法进行一步预测，前 1000 个数据用作拟合，后 100 个数据用作测试。

1. 正则化参数固定的情况

下面用 Henon 映射这一低维混沌系统来说明自适应局部线性化法(此处仅自适应确定嵌入维数，正则化参数仍取固定值)的预测能力强于一般局部线性化方法，并研究噪声对线性模型预测效果的影响。图 5.12 和图 5.13 分别给出了无噪声和有噪声情况下正规化均方根误差 NRMSE 对不同的 m 值的变化情况，其中 $\tau=3$，对主成分回归 PCR，对应于 Henon 吸引子的拓扑维 $d=1.25$，取正则化参数 $q=2$；对岭回归 RR，取 $\mu=\hat{\sigma}^2$，相应的正规化均方根误差 NRMSE 值见表 5.7。

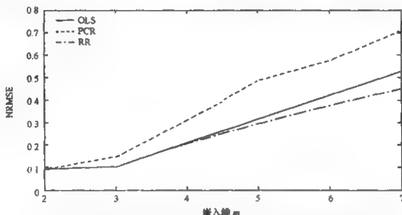


图 5.12　$K=15$ 时无噪声 Henon 映射的一般局部线性化法预测的误差

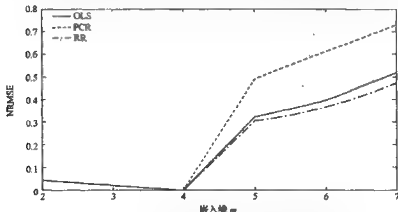


图 5.13　无噪声 Henon 映射的自适应局部线性化法预测的误差

表 5.7　无噪声 Henon 映射的正规化均方根误差 NRMSE

方法		$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$
一般方法	OLS	0.0921	0.1031	0.2970	0.3170	0.4203	0.5286
	PCR	0.0921	0.1451	0.3072	0.4846	0.5772	0.7073
	RR	0.0927	0.1031	0.2044	0.2989	0.3746	0.4499
自适应法	OLS	0.0411	0.0142	0	0.3227	0.3944	0.5179
	PCR	0.0411	0.0157	0	0.4890	0.6077	0.7334
	RR	0.0407	0.0143	0	0.3059	0.3641	0.4700

比较图 5.12 与图 5.13(或从表 5.7 中)可以看出，对无噪声的 Henon 映射，采用自适应局部线性化方法后，结果明显好于一般局部线性化方法，并且无论是用一般局部线性化方法还是用自适应局部线性化方法，OLS 的 NRMSE 均小于 PCR 的 NRMSE，这是由于正则化过滤掉了有用的预测信息。值得注意的是，对较大的 m 值，由自适应局部线性化法计算 PCR 的 NRMSE，其增长有加快的趋势，这是由于在每个目标点处，经过自适应选择 m 后，数据矩阵 X 的奇异性已大大降低了，对 PCR，如果再取较小的 q 值(此处 $q=2$)，就会过滤掉较多有用的预测信息，从而导致 NRMSE 增加，并且 m 越大，过滤掉的有用信息越多，NRMSE 增加得就越快。对 RR，有两个值变大了，也有改进的必要，后面将结合对含噪声 Henon 映射的分析，研究如何进一步减少它们的 NRMSE。由图 5.13 还可以看出，$m=4$ 时获得最好的预测结果，这与文献[3]给出的标准 $m \geqslant 2d+1$ 一致。

对有噪声的数据，使用一般局部线性化方法预测时，OLS 解随着 m 的增加变得更不稳定，并且预测能力退化了，特别当 m 接近 K 时，误差变得非常大(图 5.14、图 5.15 及表 5.8)。此时 RR 预测也失效了，这是由于低估了残差方差，因此

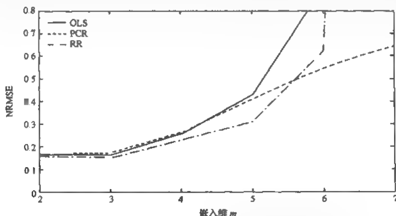


图 5.14　$K=8$ 时有噪声 Henon 映射的一般局部线性化法预测的误差

RR 方法中被噪声干扰较多的方向没有进行充分压缩。从表 5.8 中还可以看出，PCR 预测较稳定。使用自适应局部线性化法后，OLS 及 RR 的预测效果得到大大改善，特别当 m 接近 K 时，改善的效果非常明显，而 PCR 基于同样的原因大体保持不变，且略有退化。值得注意的是，使用自适应局部线性化法后，OLS 达到了与正则化相同的预测效果。

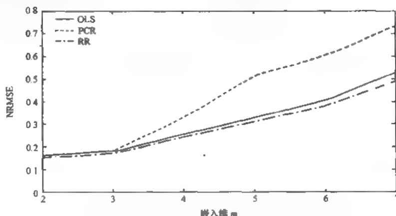


图 5.15　有噪声 Henon 映射的自适应局部线性化法预测的误差

表 5.8　有噪声 Henon 映射的正规化均方根误差 NRMSE

方法		$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$
一般方法	OLS	0.1659	0.1629	0.2538	0.4272	0.8933	6.4802
	PCR	0.1659	0.1681	0.2589	0.4059	0.5499	0.6423
	RR	0.1577	0.1517	0.2270	0.3062	0.6212	6.4802
自适应法	OLS	0.1597	0.1806	0.2534	0.3275	0.4049	0.5302
	PCR	0.1597	0.1821	0.3267	0.5100	0.6027	0.7339
	RR	0.1507	0.1704	0.2432	0.3115	0.3799	0.4889

2. 正则化参数自适应选取的情况

在表 5.7 及表 5.8 中，使用自适应局部线性化法后 PCR 及 RR 的 NRMSE 还没有达到非常满意的结果，这是由于正则化参数取的是固定值。为了进一步降低 PCR 及 RR 的误差，对不同的目标点，正则化参数 q 与 μ 应该取不同的值。首先讨论参数 q 如何自适应取值，经自适应选取 m 后，数据矩阵 X 的奇异性大大降低了。

换句话说，X 没有非常小的奇异值，所以 q 不应该取较小的值，考虑让它与矩阵 X 的条件数 γ 联系起来。文献[164]从实际应用角度给出了根据条件数 γ 判断矩阵 X 的列向量间多重共线性程度的标准：当 $\gamma<10$ 时，认为多重共线性的程度很小，当 $10<\gamma<100$ 时，则认为存在中等程度或较强的多重共线性，当 $\gamma>100$ 时，则认为存在严重的多重共线性。另外，当矩阵 X 不存在多重共线关系时，OLS 估计是最好的，即不应该在估计式 $\hat{b}_{\mathrm{OLS}}=\sum_{i=1}^{r}\frac{(u_i,y)}{\sigma_i}v_i$ 中截掉后面一些项。也就是说，此时应取 $q=r$。根据这一想法，在 X 的奇异值 $\sigma_1\geqslant\cdots\geqslant\sigma_r>0$ 中寻找满足 $\sigma_1/\sigma_i<10(i=1,2,\cdots,r)$ 且下标最大的数作为 q 的值(这样的 q 值一定存在，因为必有 $\sigma_1/\sigma_1=1<10$)。

下面再来讨论参数 μ 的选取。X 的列向量间的多重共线性程度越高(此时条件数 γ 越大)，就要求采用越强的正则化，也就是要对 OLS 估计进行更多的压缩。在实际问题中，往往由于残差的方差非常小，即 $\mu=\hat{\sigma}^2$ 接近于零，导致 RR 并没有对 OLS 估计进行充分的压缩，致使 RR 估计变得不理想，如图 5.15 所示。为了改变这种情况，考虑用

$$\mu=\gamma\hat{\sigma}^2 \tag{5.106}$$

取代 $\mu=\hat{\sigma}^2$，此时 RR 估计的过滤因子为

$$\lambda_i=\frac{\sigma_i^2}{\sigma_i^2+\gamma\hat{\sigma}^2} \tag{5.107}$$

当多重共线性程度越高时，条件数 γ 越大，过滤因子 λ_i 就越小，此时对 OLS 估计的压缩就越多。

以上给出了正则化参数 q 与 μ 的自适应取值方法，为了检验其合理性，对图 5.13 及图 5.15 中所用的数据采用所给方法再计算一遍，结果分别如图 5.16 和

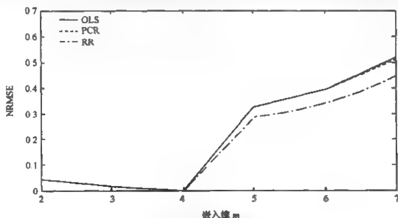


图 5.16　无噪声 Henon 映射自适应选取参数 q 与 μ 后再自适应局部线性化法预测的误差

图 5.17及表 5.9 和表 5.10 所示。

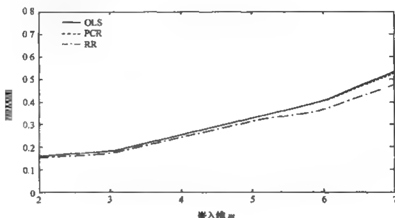


图 5.17　有噪声 Henon 映射自适应选取参数 q 与 μ 后再自适应局部线性化法预测的误差

为了方便与表 5.7 中用取固定的 q 与 μ 值所得到的结果对比，把两次计算出的 NRMSE 放入表 5.9 中。

表 5.9　自适应取 q 与 μ 前后无噪声 Henon 映射的正规化均方根误差 NRMSE

q 与 μ	方法	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$
取固定值	OLS	0.0411	0.0142	0	0.3227	0.3944	0.5179
	PCR	0.0411	0.0157	0	0.4890	0.6077	0.7334
	RR	0.0407	0.0143	0	0.3059	0.3641	0.4700
自适应取值	OLS	0.0411	0.0142	0	0.3227	0.3944	0.5179
	PCR	0.0411	0.0142	0	0.3264	0.3918	0.5132
	RR	0.0406	0.0142	0	0.2874	0.3417	0.4488

表 5.10　自适应取 q 与 μ 前后有噪声 Henon 映射的正规化均方根误差 NRMSE

q 与 μ	方法	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$
取固定值	OLS	0.1597	0.1806	0.2534	0.3275	0.4049	0.5302
	PCR	0.1597	0.1821	0.3267	0.5100	0.6027	0.7339
	RR	0.1507	0.1704	0.2432	0.3115	0.3799	0.4889
自适应取值	OLS	0.1597	0.1806	0.2534	0.3275	0.4049	0.5302
	PCR	0.1591	0.1796	0.2487	0.3196	0.4091	0.5185
	RR	0.1541	0.1709	0.2360	0.3040	0.3636	0.4780

在表 5.9 中，由于自适应选取参数 q 与 μ 对 OLS 没有影响，故 OLS 的 NRMSE 是相同的。对于 PCR 与 RR，自适应选取正则化参数 q 与 μ 以后，它们的 NRMSE 显著降低了。说明该方法对无噪声数据的效果非常好。再来看该方法对有噪声的数据是否有效，结果如表 5.10 所示。

自适应选取正则化参数 q 与 μ 以后，PCR 的 NRMSE 明显降低了，而对于 RR，多数 NRMSE 值减小了，少部分值却变大了，但减小的幅度较明显而增大的幅度不明显。经过数值模拟，说明所给出的自适应选取正则化参数 q 与 μ 的方法是有效的。

由表 5.7 及表 5.8 可以看出，自适应局部线性化法比一般局部线性化法给出了更好的预测结果，特别是当对 PCR 的参数 q 及 RR 的参数 μ 采用自适应法选择后，情况更是如此（表 5.9 及表 5.10）。当 m 接近 K 时，一般局部线性化方法的误差会非常大（图 5.14），而自适应局部线性化法则表现出良好的稳健性，其结果明显优于一般局部线性化方法，它克服了因参数 K 的选择不当而造成的影响。当观测数据中加入测量噪声时，所有方法的预测能力都下降了，但自适应局部线性化法比一般局部线性化方法具有更强的抗噪性。由于实测数据中总是含有噪声的，所以自适应局部线性化法在实际问题中具有明显的优势。

第六章　非线性时间序列分析法在证券市场中的应用

证券市场是国家经济运行的核心，探求证券市场的变化规律、进行有效的金融管理、提高金融投资效率是各国政府与投资机构追求的目标，对证券市场的各种时间序列进行分析、预测和控制是整个证券市场活动的重要工作。我国证券市场经过20几年的高速发展，已经形成了相当规模的市场，证券市场是一个受多种因素影响的、庞大的系统，具有非常复杂的运动规律。取自相同时间段的多种证券时间序列，构成多变量证券时间序列。这些证券时间序列则是证券市场的外在表现形式。"本质决定现象，现象反映本质"，因此证券时间序列中必定蕴含了市场复杂系统许多客观规律信息。从中寻找出各种信息，更好地认识、掌握、并利用其规律无疑对证券市场的预测、决策与风险管理活动具有特别重要的意义。

6.1　基于单变量时间序列的证券市场非线性性和确定性检验

长期以来，资本市场理论为线性范式所主宰，即有效市场假设，它假定投资者是理性的、有秩序的、有条理的，它把数学约化到简单的线性微分方程有关的投资行为模型，认为证券市场是一个随机游走过程，价格变化函数的概率分布服从正态分布。然而，证券市场却不是秩序的或简单的，它们既混乱又复杂。E. 彼得斯在他的《资本市场的混沌和秩序》和《分形市场分形——将混沌理论应用到投资与经济理论上》等书中提出了一种资本市场的新范式，在资本市场中引入了混沌理论[165,166]。正如E. 彼得斯说的那样要证明资本市场是"混沌的"说法几乎已经（并将继续）成为笑话，但混沌现象必然发生在非线性系统内，因此研究证券市场的非线性性和确定性是一件有意义且非常重要的事情。

6.1.1　样本数据及平稳化处理

选择以下单变量证券市场指数时间序列作为研究对象：上海证券市场综合指数（上证，1997.1.2～2002.12.27 数据）、香港恒生指数（恒指，1997.1.2～2002.12.27 数据）、台湾加权指数（台湾，1997.1.2～2002.12.27 数据）、美国标准普尔 500 指数（普尔，1997.1.2～2002.12.27 数据）。表 6.1 是四组证券市场指数

时间序列的统计分析结果。

表 6.1　各证券市场指数时间序列的统计数据

统计项	上证指数	恒生指数	台湾加权指数	普尔 500 指数
均值	1536	12 468	6871.96	778.68
标准差	336.74	2597.75	1650.94	378.51
偏度	0.3444	0.1516	−0.0287	0.4217
峰度	−0.9785	−0.8671	−1.0416	−1.3191
Jarque Bera	84.8333	51	59.72	323.83

非线性时间序列的很多处理方法要求时间序列是平稳的，但证券市场指数时间序列不能满足平稳性的要求，为此，在进行进一步分析之前首先需要对这些时间序列进行平稳处理。为了保证处理后的时间序列具有实际意义，通常采用差分和去除趋势的方法进行平稳处理[165~168]。目前，在证券时间序列分析中广泛使用的方法有以下两种：

1）收益率(rate of return，ROR)平稳化方法。设$\{P_n\}$为指数序列，则收益率平稳化后的时间序列为

$$x_n = \ln P_n - \ln P_{n-1} \tag{6.1}$$

2）对数线性去趋势(log-linear de-trending，LLD)平稳化方法。以$\ln P_n$为因变量，n为自变量，做以下对数线性回归

$$\ln P_n = a + bn \tag{6.2}$$

得对数线性去趋势平稳化后的时间序列为

$$x_n = \ln P_n - (a + bn) \tag{6.3}$$

收益率平稳化方法处理后时间序列的统计数据如表 6.2 所示，对数线性去趋势平稳化方法处理后时间序列的统计数据如表 6.3 所示。

表 6.2　各证券市场指数时间序列经收益率平稳化后的统计数据

统计项	上证指数	恒生指数	台湾加权	标准普尔 500
均值	0	−0.0001	0	−0.0139
标准差	0.1507	0.2132	0.1653	0.181
偏度	0.0757	−0.2353	0.2365	−0.0925
峰度	−1.0084	−0.6272	0.0311	0.1567
Jarque Bera	61.6092	37.2768	12.3254	7.9804

表 6.3　各证券市场指数时间序列经对数线性去趋势平稳化后的统计数据

统计项	上证指数	恒生指数	台湾加权	标准普尔 500
均值	0.0004	0	−0.0003	0.0004
标准差	0.0164	0.0208	0.0191	0.0105
偏度	0.0519	0.513	0.075	−0.0307
峰度	6.086	8.9405	1.7945	3.7578
Jarque Bera	2193.7	4909.8	169.063	1918.1

图 6.1 是上海证券市场综合指数时间序列通过收益率(ROR)平稳化和对数线性去趋势(LLD)平稳化后的时间历程图，从图 6.1 中可以发现，平稳化后的时间序列中，对数线性去趋势平稳化方法产生的时间序列的平稳性要比收益率平稳化方法产生的时间序列的平稳性差，前者的标准差是后者的 9.19 倍。

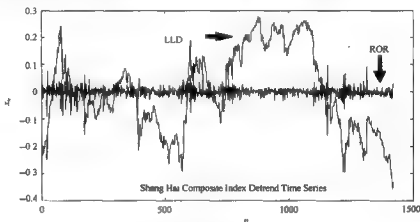


图 6.1　上证综合指数两种不同方法平稳化后的时间历程图

图 6.2 是上海证券市场综合指数时间序列通过收益率平稳化后的时间序列的直方图。通过图 6.2 很容易发现，上海证券市场综合指数的分布具有“尖峰胖尾”的特点，不符合传统的有效市场假设，香港恒生指数、台湾加权指数和美国标准普尔 500 指数的情况类似。

6.1.2　证券市场的非线性性检验

采用 2.5 节的替代数据法对上述平稳化后的时间序列进行非线性检验，其中零假设为原时间序列来自具常系数的高斯线性随机过程经过单调静态非线性变换，利用 AAFT 算法生产替代时间序列，检验结果如表 6.4 所示。

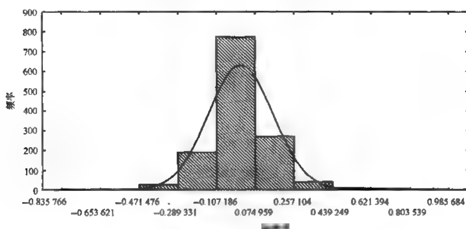


图 6.2　上证综合指数通过收益率平稳化后的直方图

表 6.4　各证券市场指数序列非线性性的检验结果

收益率时间序列的替代法检验				
判别统计量	上证指数	恒生指数	台湾加权	标准普尔 500
最大 Lyapunov 指数	√	√	√	√
关联维数	√	√	√	√
一步预测误差	√	√	√	√
对数线性去趋势时间序列的替代法检验				
判别统计量	上证指数	恒生指数	台湾加权	标准普尔 500
最大 Lyapunov 指数	×	√	√	×
关联维数	√	×	×	√
一步预测误差	√	×	×	√

注：√表示能有效拒绝零假设，×表示不能有效拒绝零假设。

从表 6.4 中可以看出，上海证券市场综合指数、香港恒生指数、台湾加权指数和美国标准普尔 500 指数时间序列经对数线性去趋势平稳化后的时间序列，在采用最大 Lyapunov 指数、关联维数和一步预测误差作为判别统计量时，未能很好地拒绝零假设；当原指数时间序列经收益率平稳化后的时间序列，在采用最大 Lyapunov 指数、关联维数和一步预测误差作为判别统计量时，能有效地拒绝零假设。进一步利用 R/S 分析法和 BDS 检验方法可以判断出上述四个证券市场均具有一定程度的非线性，详细内容可见文献[169,170]。

6.1.3　证券市场的确定性检验

使用递归图可以通过视觉直接审视被检测时间序列是随机无序结构的还是有可识别模式的。如果可以找出特定的模式则可以进一步假设该时间序列具有一定的确定性，该时间序列中的确定性可以通过某些特征值、递归图中点的分布规律等进一步检测获得。如果该时间序列是随机的，则它递归图中的点应均匀分布在整个图形中，从中无法识别出固定的模式。如果时间序列中具有确定性，则递归图中点的分布具有规律性。例如，在对角线周围有平行与对角线的线段，则表明该时间序列具有一定的预测能力。通常递归图主要用于对时间序列进行定性分析。

图 6.3 和图 6.4 分别是 6.1.1 节中的上海综合指数经收益率平稳化和经对数线性去趋势平稳化后的时间序列的递归图，两个图中均可以在对角线周围发现明显的平行线段。6.1.1 节中的其他三组证券时间序列经平稳化后均可以发现类似现象。据此可以认为上海证券市场综合指数、香港恒生指数、台湾加权指数和美国标准普尔 500 指数时间序列经平稳化后具有较明显的确定性。

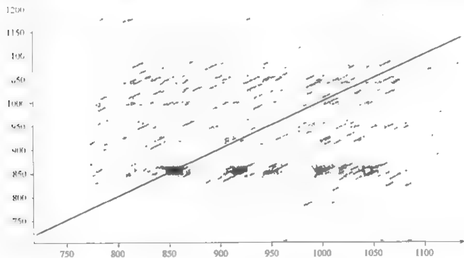


图 6.3　上证指数经收益率平稳化后的递归图

图 6.3 和图 6.4 所示的递归图中间有很多复杂的结构通过视觉很难被辨识，为此进一步引入与递归图紧密相关的特征量，比较常用的特征量有以下两个：

(1) 递归点比例(REC)

递归点比例(REC)的计算公式为

$$\mathrm{REC} = \frac{N_r}{N_t} \tag{6.4}$$

式中，N_r 是所有递归点(递归图中的黑点)；N_t 是所有可能的递归点的个数。

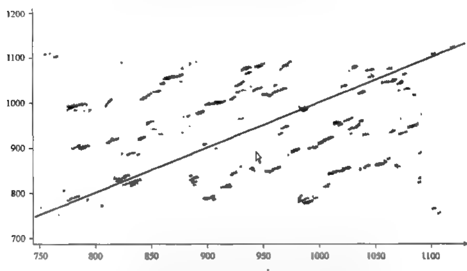


图 6.4　上证指数经对数线性去趋势平稳化后的递归图

(2) 确定性比例(DET)

确定性比例(DET)的计算公式为

$$\mathrm{DET}=\frac{N_l}{N_r} \tag{6.5}$$

式中，N_l 是包含在与对角线平行线段中的递归点个数。

表 6.5 和表 6.6 分别给出了 6.1.1 节中所研究的四个证券时间序列经收益率平稳化和经对数线性去趋势平稳化处理后时间序列递归图的 REC 和 DET，为了便于对比，表中同时提供了该时间序列经随机变换方式和 AAFT 方法产生的替代时间序列对应的 REC 和 DET。

从表 6.5 和表 6.6 中可以看出，对于 REC 和 DET 而言，经随机方法和 AAFT 方法生成的替代时间序列比经收益率平稳化和经对数线性去趋势平稳化处理后的时间序列的值明显下降。根据 REC 和 DET 的定义，表明上述四个证券时间序列对应的复杂系统具有较明显的确定性。

表 6.5　原始和替代时间序列的 REC 计算结果对比

项目	收益率平稳化后的 REC			对数线性去趋势平稳化后的 REC		
	原始	随机	AAFT	原始	随机	AAFT
上证	0.036	0.001	0.003	0.114	0.001	0.002
恒指	0.191	0.021	0.032	0.162	0.001	0.124
台湾	0.012	0.001	0.011	0.257	0.001	0.191
普尔	0.028	0.003	0.004	0.141	0.001	0.231

表 6.6　原始和替代时间序列的 DET 计算结果对比

项目	收益率平稳化后的 DET			对数线性去趋势平稳化后的 DET		
	原始	随机	AAFT	原始	随机	AAFT
上证	0.461	0.001	0.001	0.612	0.001	0.001
恒指	0.511	0.001	0.001	0.823	0.001	0.701
台湾	0.012	0.011	0.011	0.905	0.001	0.801
普尔	0.521	0.001	0.001	0.711	0.001	0.691

6.2　基于多变量时间序列的证券市场非线性性检验

证券市场中常常可以获得同时测量的多组时间序列构成多变量时间序列。例如，证券市场的分类指数时间序列、分类成交量时间序列等。本节将利用第三章的基于多变量时间序列的系统非线性性检验方法分析证券市场的非线性特征。

6.2.1　样本数据及平稳化处理

选取上海证券市场分类指数的每日收盘指数作为研究对象，具体样本为 1997 年 1 月 2 日到 2002 年 4 月 30 日的工业类 $\{P_{1,n}\}$、商业类 $\{P_{2,n}\}$、地产类 $\{P_{3,n}\}$ 和公用事业类 $\{P_{4,n}\}$ 指数，构成四个变量的时间序列 $\{P_{l,n}\}l=1,2,3,4$，其时间历程图如图 6.5 所示。

由于证券市场指数时间序列有随经济增长的趋势，因此必须先去除这种趋势，使时间序列平稳化。类似于单变量的情况，用以下两种去趋势方法处理。

(1) 收益率平稳化方法

设 $\{P_{l,n}\}$，$l=1,2,3,4$ 为指数序列，则收益率平稳化后的序列为

$$x_{l,n} = \ln P_{l,n} - \ln P_{l,n-1},\quad l = 1,2,3,4 \tag{6.6}$$

则由工业类 $\{P_{1,n}\}$、商业类 $\{P_{2,n}\}$、地产类 $\{P_{3,n}\}$ 和公用事业类 $\{P_{4,n}\}$ 指数构成的四变量时间序列通过收益率平稳化后的时间序列历程图如图 6.6 所示。

(2) 对数线性去趋势平稳化方法

以 $\ln P_{l,n}$ 为因变量，n 为自变量，做以下对数线性回归

$$\ln P_{l,n} = a_l + b_l n \tag{6.7}$$

得对数线性去趋势平稳化后的序列为

$$x_{l,n} = \ln P_{l,n} - (a_l + b_l n),\quad l = 1,2,3,4 \tag{6.8}$$

则由工业类 $\{P_{1,n}\}$、商业类 $\{P_{2,n}\}$、地产类 $\{P_{3,n}\}$ 和公用事业类 $\{P_{4,n}\}$ 指数构成的四变量时间序列通过对数线性去趋势平稳化后的时间序列历程图如图 6.7 所示。

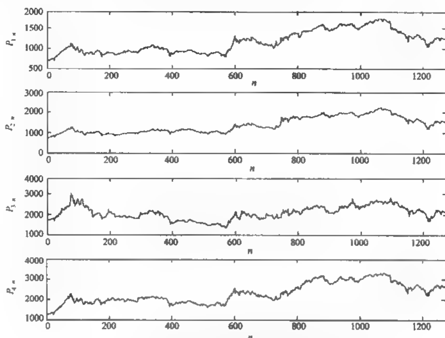


图 6.5　上海证券市场的分类指数序列时间历程图

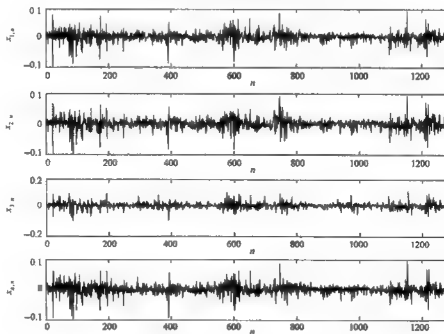


图 6.6　收益率平稳化后的分类指数序列时间历程图

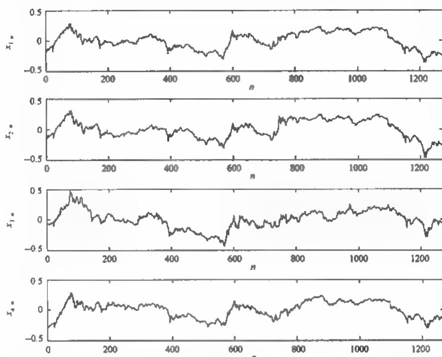


图 6.7　对数线性去趋势平稳化后的分类指数序列时间历程图

对上述去趋势后的时间序列中的每个变量时间序列标准化，即得到各变量均值为零，方差为 1 的多变量时间序列。

6.2.2　证券市场的非线性性检验

零假设 1：时间序列来自具常系数的高斯线性随机过程。

利用 2.5.1 节中随机相位化 Fourier 变换(FT)方法产生 19 组替代多变量时间序列，对经过收益率平稳化后的时间序列，当 $r=0.3$ 时，经计算，线性冗余的显著性检验统计量 $\mathrm{DIFF}L(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$关于时间延迟间隔 τ 的函数如图 6.8(a)所示，广义冗余显著性检验统计量 $\mathrm{DIFF}R_q(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$关于时间延迟间隔 τ 的函数如图 6.8(b)($q-1$)、图 6.8(c)($q=2$)和图 6.8(d)($q=3$)所示。

线性边缘冗余 $\lambda(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$与广义边缘冗余 $\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$的显著性检验统计量 $\mathrm{DIFF}\lambda(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$和 $\mathrm{DIFF}\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$的计算结果如图 6.9 所示。

对经过对数线性去趋势平稳化后的时间序列，当 $r=0.1$ 时，经计算，线性冗余的显著性检验统计量 $\mathrm{DIFF}L(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$关于时间延迟间隔 τ 的函数如

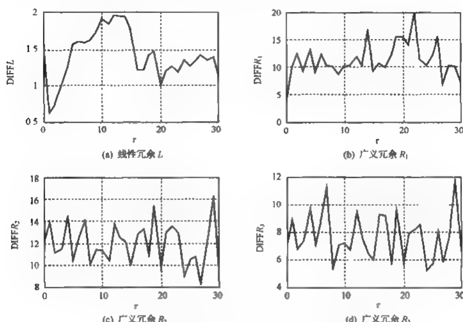


图 6.8　收益率分类指数序列在零假设 1 下的显著性检验统计量

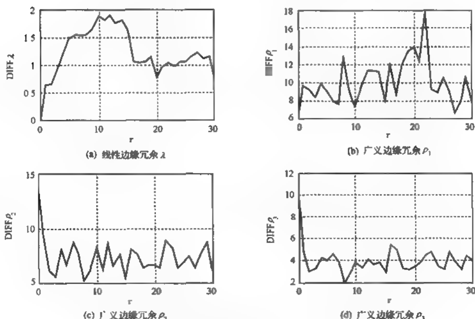


图 6.9　收益率分类指数序列在零假设 1 下的显著性检验统计量

图 6.10(a)所示，广义冗余显著性检验统计量 $\mathrm{DIFF}R_q(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$关于时间延迟间隔 τ 的函数如图 6.10(b)($q=1$)、图 6.10(c)($q=2$)和图 6.10(d)($q=3$)所示。

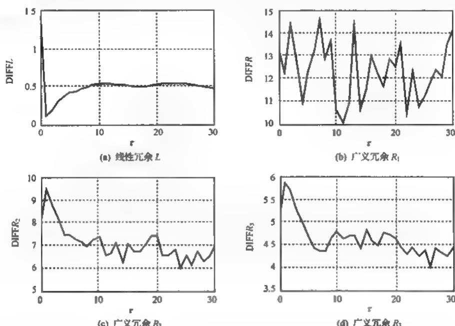


图 6.10　对数线性去趋势后的分类指数序列在零假设 1 下的显著性检验统计量

线性边缘冗余 $\lambda(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$与广义边缘冗余 $\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$的显著性检验统计量 $\mathrm{DIFF}\lambda(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$和 $\mathrm{DIFF}\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$的计算结果如图 6.11 所示。

综上讨论可见，无论是利用收益率平稳化还是利用对数线性去趋势平稳化方法，线性冗余的显著性检验统计量的数值都小于 1.96，这表明原实测时间序列与替代时间序列的线性依赖结构是一样的，即替代时间序列在技术上是好的，不应该产生虚假的检验结果。但非线性冗余的显著性检验统计量的数值较大(>1.96)，则拒绝零假设 1，因此可以认为上海证券市场分类指数时间序列不是线性高斯随机序列，而是具有非线性特性。

零假设 2：时间序列来自具常系数的高斯线性随机过程经过单调静态非线性变换。

利用 2.5.1 节中迭代的幅度调节 Fourier 变换(IAAFT)方法生产 19 组替代多变量时间序列，由于此时采用的是 Fourier 递归型快速算法，要求时间序列长度满足 $N=2^i3^j5^k(i,j,k=0,1,\cdots)$，因此从原 1273 个交易日中截取每一变量长度为 1250 的四变量分类指数时间序列。

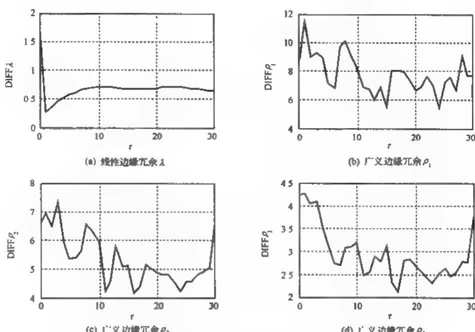


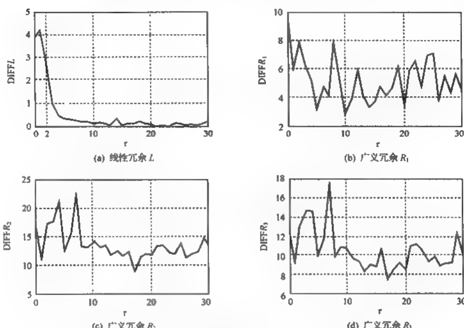
图 6.11 对数线性去趋势后的分类指数序列在零假设 1 下的显著性检验统计量

对经过收益率平稳化后的时间序列，当 $r=0.4$ 时，经计算，线性冗余的显著性检验统计量 $\mathrm{DIFF}L(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$关于时间延迟间隔 τ 的函数如图 6.12(a)所示，广义冗余显著性检验统计量 $\mathrm{DIFF}R_q(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$关于时间延迟间隔 τ 的函数如图 6.12(b)($q=1$)、图 6.12(c)($q=2$)和图 6.12(d)($q=3$)所示。

可见，线性冗余的显著性检验统计量的数值除 $\tau=1,2$ 外几乎都很小；$\tau=1,2$ 时较大，这可能是由于时间间隔延迟和样本不够，而非线性冗余的显著性检验统计量的数值都大于 1.96，这样可以拒绝零假设 2。因此可以认为，上海证券市场分类指数序列不是具常系数的高斯线性随机过程经过单调静态非线性变换获得的，而是具有非线性特性。

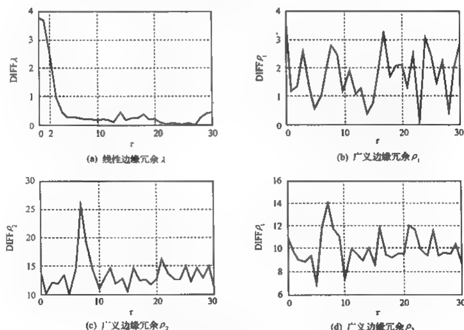
线性边缘冗余 $\lambda(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$与广义边缘冗余 $\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$的显著性检验统计量 $\mathrm{DIFF}\lambda(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$和 $\mathrm{DIFF}\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$的计算结果如图 6.13 所示。

可以发现线性边缘冗余显著性检验统计量的数值除了 $\tau=1,2$ 外都很小，$q=1$ 时的非线性冗余显著性检验统计量却很小不能拒绝零假设。而下面可以看到对数线性去趋势平稳化方法不存在这样的问题，这或许正如文献[166]所说的收益率的 $AR(1)$ 残差分析可能掩盖了显著性信息，收益率平稳化在样本不足而采用多变量



(a) 线性冗余 L　(b) 广义冗余 R_1　(c) 广义冗余 R_2　(d) 广义冗余 R_3

图 6.12　收益率分类指数序列在零假设 2 下的显著性检验统计量



(a) 线性边缘冗余 λ　(b) 广义边缘冗余 ρ_1　(c) 广义边缘冗余 ρ_2　(d) 广义边缘冗余 ρ_3

图 6.13　收益率分类指数序列在零假设 2 下的显著性检验统计量

IAAFT 替代数据法情况下容易掩盖显著性信息，于是影响了 $q=1$ 时的非线性边缘冗余的显著性检验统计量 $\mathrm{DIFF}R_q(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$。$q=2,3$ 时的非线性冗余显著性检验统计量可以有效地拒绝零假设 2。

对经过对数线性去趋势平稳化后的时间序列，当 $r=0.1$ 时，经计算，线性冗余的显著性检验统计量 $\mathrm{DIFF}L(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$ 关于时间延迟间隔 τ 的函数如图 6.14(a)所示，广义冗余显著性检验统计量 $\mathrm{DIFF}R_q(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$ 关于时间延迟间隔 τ 的函数如图 6.14(b)($q=1$)、图 6.14(c)($q=2$)和图 6.14(d)($q=3$)所示。

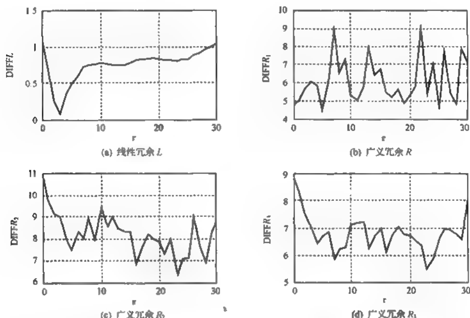


图 6.14　对数线性去趋势后分类指数序列在零假设 2 下的显著性检验统计量

线性边缘冗余 $\lambda(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$ 与广义边缘冗余 $\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$ 的显著性检验统计量 $\mathrm{DIFF}\lambda(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$ 和 $\mathrm{DIFF}\rho_q(x_{1,n};x_{2,n+\tau};x_{3,n};x_{4,n})$ 的计算结果如图 6.15 所示。

可见，线性冗余的显著性检验统计量的数值都小于 1.5，这表明原实测时间序列与替代时间序列的线性依赖结构是一样的，即替代时间序列在技术上是好的，不应该产生虚假的检验结果，但非线性冗余的显著性检验统计量的数值较大(>1.96)，则拒绝零假设 2。因此可以认为，上海证券市场分类指数时间序列不是来自具常系数的高斯线性随机过程经过单调静态非线性变换，而是具有非线性特性。

上面利用两种多变量时间序列替代数据生成技术，即随机相位化的 Fourier

变换(FT)和迭代的幅度调节 Fourier 变换(IAAFT)检验了上海证券市场分类指数时间序列的非线性。以线性冗余统计量检验生成的替代数据的好坏,而当线性冗余的显著性检验统计量保证了生成的替代时间序列在技术上是好的时,用非线性冗余统计量检验原时间序列的非线性性方能保证检验结果的有效性。采用线性冗余和非线性冗余两种统计量,更能避免虚假检验,且非线性冗余可用广义关联积分计算,具有计算简便、速度快等优点。

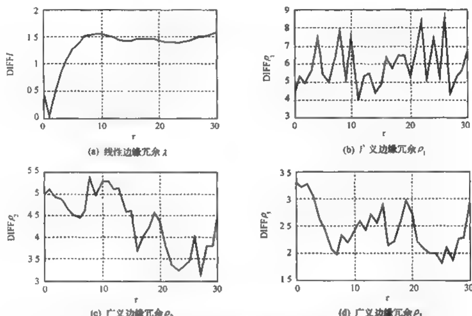


图 6.15　对数线性去趋势后分类指数序列在零假设 2 下的显著性检验统计量

计算结果表明这种方法是可靠的,检测到上海证券市场分类指数时间序列具有非线性特性,它不是线性高斯随机序列,也不是高斯随机序列,这一结论与用 R/S分析法得到的结果相吻合[171,172],表明了上海证券市场确实不是线性系统,对传统的市场有效性假设提出了疑问。

6.3　上海证券市场单变量指数序列的预测研究

传统的证券市场指数序列的预测是基于有效市场假设的,该假设认为价格反映了所有的公开信息,价格只有在收到新的信息时才会变动,因此证券市场是服从正态分布的。但大量研究表明实际情况不是这样,证券市场是非线性的,具有某种确定性甚至混沌特征,而不是随机的,因此基于非线性时间序列分析方法对证券市场的指数序列进行短期预测是可能的和可行的。

6.3.1　样本数据及相空间重构

取 1997 年 1 月 2 日到 2002 年 4 月 30 日的上海证券市场综合指数 P_n 作为研究对象，其时间历程图如图 6.16 所示。

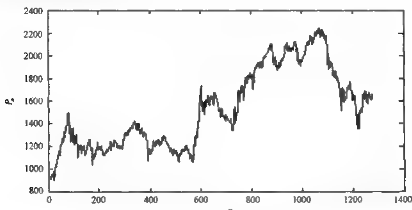


图 6.16　上海证券市场综合指数的时间历程图

为了消除证券市场指数时间序列有随经济增长的趋势，利用对数线性去趋势平稳化方法。首先以 $\ln P_n$ 为因变量，n 为自变量，作一阶线性回归，即求直线 $a+bn$，使其到各点 $\ln P_n$ 的距离平方和最小，由此可求得 a、b 的值。然后作残差，可得时间序列 $\{x_n\}$，即

$$x_n = \ln P_n - (a + bn) \tag{6.9}$$

采用对数线性去趋势平稳化方法得时间序列 $\{x_n\}_{n=1}^{N}$ 的时间历程图如图 6.17 所示。

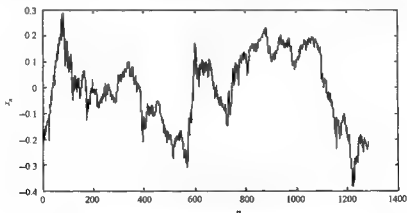


图 6.17　上海证券市场综合指数经对数线性去趋势后的时间历程图

下面对经过对数线性去趋势后的时间序列进行相空间重构。首先，用平均互信息法求延迟间隔 τ，计算结果如图 6.18 所示。

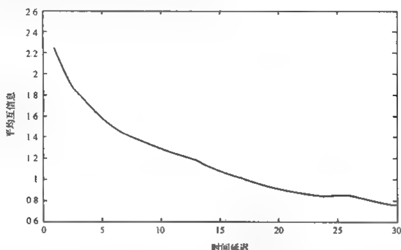


图 6.18　平均互信息法关于延迟间隔 τ 的变化情况

由图 6.18 可以看出，平均互信息曲线在 $\tau=24$ 处取得第一个极小值，所以对数线性去趋势后的时间序列的延迟时间取为 $\tau=24$。

然后利用虚假最近邻点法求嵌入维数 m，重构后计算的虚假最近邻点百分比对维数的变化情况如图 6.19 所示。

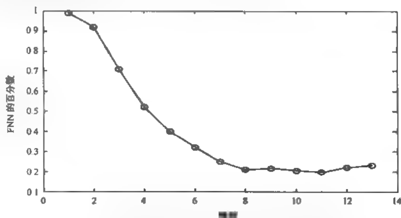


图 6.19　虚假最近邻点百分比对维数的变化情况

由于股市数据中存在着噪声，所以当维数增加到足够大时，图 6.19 中虚假最近邻点的百分比达不到零。但当 $m=8$ 时，虚假最近邻点的百分比达到最小，因此

取嵌入维数约为 8。

这样可对经对数线性去趋势后的上海证券市场综合指数做以下相空间重构

$$x_n = (x_n, x_{n-\tau}, \cdots, x_{n-7\tau}) \tag{6.10}$$

6.3.2 基于正则化的自适应预测

在对经过对数线性去趋势后的上海证券市场综合指数进行预测时，把观测数据分成两个部分，其中后 100 个点用来检验预测效果。

在基于正则化方法进行上海证券市场指数时间序列预测时，采用主成分回归(PCR)时，要确定参数 q 的值，为此必须知道吸引子的分形维数。文献[173,174]均对上海证券市场综合指数的分形维数进行了研究，都给出了其分形维数在 2 和 3 之间的结论，于是这里取 $q=3$。

为了检验用虚假最近邻点法求出的嵌入维 $m=8$ 是否合适，考虑其邻近的值，即取嵌入维数分别为 5～10 进行相空间重构(τ 均取 24)，利用普通最小二乘的局部线性预测法(OLS)和正则化的局部线性预测法(PCR，RR)，计算一步预测的正规化均方根误差，结果如图 6.20 所示(对 PCR，选参数 q 时，阈值取 7)。利用自适应的局部线性预测法计算一步预测的正规化均方根误差，结果如图 6.21 所示；具体数值如表 6.7 所示。

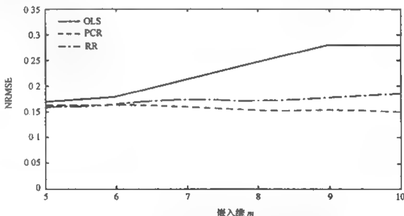


图 6.20　$K=15$ 时的一般局部线性化法预测的误差

比较图 6.20 与图 6.21(或表 6.7)可以清楚地看出，对上海证券市场综合指数进行的一步预测，自适应局部线性化方法好于一般局部线性化方法，并且由各种方法计算出的误差在 $m=8$ 处出现明显下降，特别对于自适应局部线性化法，$m=8$ 时的误差是最小的(虽然 RR 算法在 $m=9$ 处达到最小，但与 $m=8$ 处的误差相差不大)，所以用虚假最近邻点法求出的嵌入维数是比较理想的。

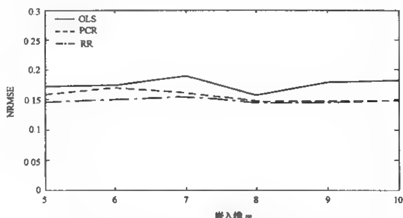


图 6.21　自适应局部线性化法预测的误差

表 6.7　上海证券市场综合指数一步预测正规化均方根误差 NRMSE

方法		$m=5$	$m=6$	$m=7$	$m=8$	$m=9$	$m=10$
一般方法	OLS	0.1676	0.1796	0.2163	0.2474	0.2769	0.2778
	PCR	0.1623	0.1657	0.1600	0.1511	0.1528	0.1506
	RR	0.1574	0.1656	0.1752	0.1700	0.1773	0.1833
自适应法	OLS	0.1721	0.1752	0.1900	0.1571	0.1781	0.1817
	PCR	0.1593	0.1697	0.1619	0.1469	0.1479	0.1479
	RR	0.1463	0.1496	0.1545	0.1445	0.1439	0.1483

取 $m=8,\tau=24$ 重构相空间，并以 RR 算法为例，分别用一般局部线性化方法和自适应局部线性化方法给出对最后 100 个点的一步预测值和实际值的比较，结果如图 6.22 和图 6.23 所示。

对比图 6.22 和图 6.23 可以很明显地看出，用自适应局部线性化方法作的一步预测，其结果与真实值比较接近，且对指数的每个高点和低点的描述和实际情况一致。用一般局部线性化法作的一步预测，不仅误差增大了，而且对细节的描述也有所减弱，表明自适应局部线性化法好于一般局部线性化法，它是一个更稳健的方法。

关于上海证券市场单变量指数时间序列的预测我们还从不同的角度进行了研究，详见文献[175～177]。

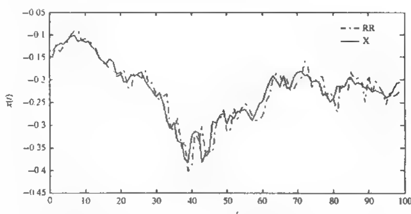


图 6.22　对数线性去趋势后上证综合指数的一般局部线性化一步预测

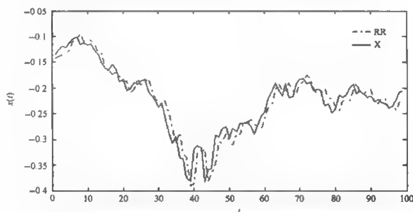


图 6.23　对数线性去趋势后的上证综合指数的自适应局部线性化一步预测

6.4　上海证券市场多变量指数序列的预测研究

目前，大量关于证券市场指数序列的预测研究都是基于单变量时间序列的，但证券市场中往往可以获取多变量时间序列，它们包含比单变量时间序列更多的信息。根据第五章的讨论，多变量时间序列的预测效果要比单变量时间序列的预测效果好，因此本节利用上海证券市场的综合指数和成交金额构成两变量时间序列，从中对综合指数进行预测研究。

6.4.1　样本数据及相空间重构

选取上海证券市场的每日收盘指数，即上证综合指数作为研究对象，样本选取1999年1月4日至2003年12月31日的上海证券市场综合股指和成交金额各1196个数据，P 表示综合股指，Q 表示成交金额指数，于是得到两个时间序列 $\{P_n\}_{n=1}^{1196}$ 和 $\{Q_n\}_{n=1}^{1196}$，如图6.24和图6.25所示。

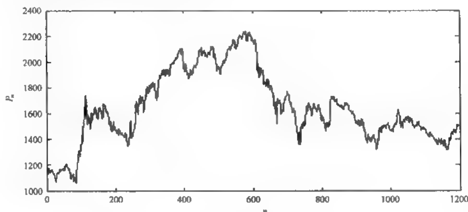


图6.24　上海证券市场综合股指序列时间历程图

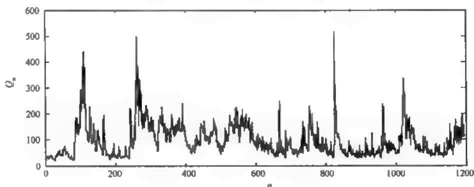


图6.25　上海证券市场成交金额指数序列时间历程图

为了使这两组时间序列有可比性，将时间序列 $\{P_n\}_{n=1}^{1196}$ 和 $\{Q_n\}_{n=1}^{1196}$ 分别进行标准化，即

$$x_{1,n} = \frac{P_n - P_{\min}}{P_{\max} - P_{\min}},\quad n = 1,2,\cdots,1196 \tag{6.11}$$

$$x_{2,n} = \frac{Q_n - Q_{\min}}{Q_{\max} - Q_{\min}}, \quad n = 1,2,\cdots,1196 \tag{6.12}$$

这样得到两个新的时间序列$\{x_{1,n}\}_{n=1}^{1196}$和$\{x_{2,n}\}_{n=1}^{1196}$，如图 6.26 和图 6.27 所示。

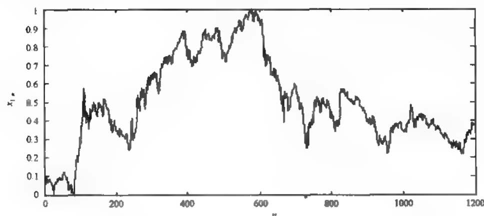


图 6.26　上海证券市场综合股指序列标准化后的时间历程图

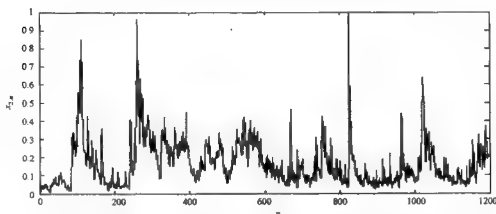


图 6.27　上海证券市场成交金额指数序列标准化后的时间历程图

文献[178]中说明了用$\{x_{1,n}\}_{n=1}^{1196}$和$\{x_{2,n}\}_{n=1}^{1196}$重构的相空间与用$\{P_n\}_{n=1}^{1196}$和$\{Q_n\}_{n=1}^{1196}$重构的相空间具有相同的最佳延迟时间间隔和嵌入维数，且它们的关联维数和 Lyapunov 指数等几何不变量也是一样的。因此，可以直接利用$\{x_{1,n}\}_{n=1}^{1196}$和$\{x_{2,n}\}_{n=1}^{1196}$来重构相空间进行预测。

为了比较预测结果，用下列两种方式重构吸引子。第一种方式是用单变量局部多项式预测法重构吸引子。由平均互信息法和 L. Y. Cao 的方法[15]，得到$\tau=34$和

$m=9$。因此重构向量为

$$x_n = \{x_{1,n}, x_{1,n-\tau}, x_{1,n-2\tau}, \cdots, x_{1,n-8\tau}\} \quad (6.13)$$

第二种方式是用多变量局部多项式预测法重构吸引子。利用 4.3.3 节提出的虚假最近邻点法确定嵌入维数算法的改进算法，得 $\tau_1=34$，$\tau_2=35$，$m_1=5$，$m_2=3$，则 $m=8$。因此重构向量为

$$x_n = \{x_{1,n}, x_{1,n-\tau_1}, x_{1,n-2\tau_1}, x_{1,n-3\tau_1}, x_{1,n-4\tau_1}, x_{2,n}, x_{2,n-\tau_2}, x_{2,n-2\tau_2}\} \quad (6.14)$$

对双变量时间序列 $\{x_{1,n}, x_{2,n}\}_{n=1}^{1196}$，前 1100 个数据用于训练样本，后 96 个数据用于预测。

6.4.2 局部多项式预测

下面利用第五章提出的多变量时间序列的局部多项式预测法对标准化后的上海证券市场综合指数(即第一个变量)进行预测。取邻点数为 $K=80$，在式(5.19)中分别取多项式阶数 $s=2$ 和预测步数 $p=1$，一步预测的均方根误差和正规化均方误差如表 6.8 所示。

表 6.8 单变量和两变量局部多项式预测的一步预测误差

方法	时间序列	均方根误差 RMSE	正规化均方误差 NMSE
局部多项式预测	单变量	0.0276	0.4452
	多变量	0.0175	0.1798

由表 6.8 知，用多变量局部多项式法预测的结果很好地预测了真实的时间序列，并且当使用相同邻点数时这种方法预测结果明显优于单变量情形，如图 6.28 所示。因此，这种多变量的局部多项式预测方法是成功的。

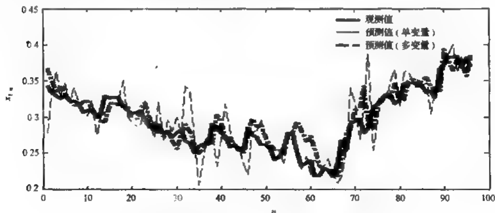


图 6.28 局部多项式预测法预测上海综合股指的最后 96 个数据

6.4.3 基于正则化的局部线性和局部多项式预测

下面利用 5.4 节提出的基于正则化的多变量时间序列的局部线性和局部多项式预测法对标准化后的上海证券市场综合指数进行预测。仍取 $\tau_1=34$，$\tau_2=25$，$m_1=5$，$m_2=3$，重构向量见式(6.14)。对双变量时间序列 $\{x_{1,n},x_{2,n}\}_{n=1}^{1196}$，前 1100 个数据用于训练样本，后 96 个数据用于预测。根据文献[173,174]对上海综合指数的分形维数进行了研究，给出了其分形维数在 2 和 3 之间的结论，故对 PCR 取正则化参数 $q=3$；对 RR，取 $\mu=\hat{\sigma}^2$。采用正规化均方根误差 NRMSE 来衡量一步预测的精度，图 6.29 表示分别利用普通最小二乘的局部线性预测法(OLS)和利用正则化的局部线性预测法(PCR，RR)对标准化后的上海证券市场综合指数最后 96 个数据进行一步预测得到的正规化均方根误差 NRMSE。

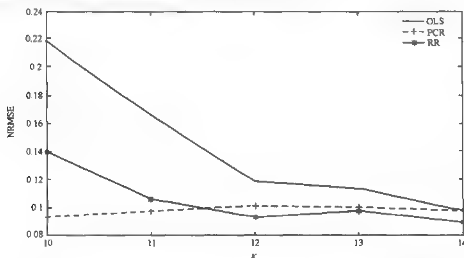


图 6.29 标准化后的上海证券市场综合指数的局部线性预测的一步预测误差

图 6.30 表示分别利用普通最小二乘的局部多项式预测法(OLS)和利用正则化的局部多项式预测法(PCR，RR)对标准化后的上证综合指数最后 96 个数据进行一步预测的正规化均方根误差 NRMSE。此时取多项式的阶数 $s=2$，则待定系数个数 $J=\sum_{i=1}^{2}C_{m_i+2}^{2}-2+1=C_7^2+C_5^2-2+1=30$。

表 6.9 表示不同邻点 K 下分别利用普通最小二乘的局部线性预测法(OLS)和利用正则化的多变量局部线性预测法(PCR，RR)对标准化后的上证综合指数最后 96 个数据进行预测的一步预测误差值比较。

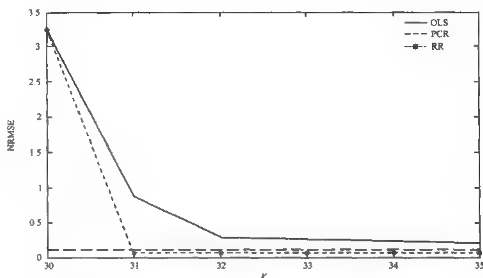


图 6.30　标准化后的上海证券市场综合指数的局部多项式预测的一步预测误差

表 6.9　上证综合股指的局部线性预测法的正规化均方根误差 NRMSE

NRMSE	K>m						K=m	K<m
	14	13	12	11	10	9	8	7
OLS	0.0981	0.1136	0.1189	0.1708	0.2184	7.3450	—	—
PCR	0.0980	0.1003	0.0888	0.1015	0.0933	0.0960	0.0952	0.0968
RR	0.0894	0.0975	0.0939	0.0935	0.1403	—	—	

表 6.10 表示不同邻点 K 下分别利用普通最小二乘的局部多项式预测法(OLS)和利用正则化的局部多项式预测法(PCR，RR)对标准化后的上海证券市场综合指数最后 96 个数据进行预测的一步预测误差值比较。

表 6.10　上证综合股指的局部多项式预测法的正规化均方根误差 NRMSE

NRMSE	K≥J						K<J	
	35	34	33	32	31	30	29	28
OLS	0.2239	0.2247	0.2436	0.3009	0.8708	3.2265	8.2326	2.1346
PCR	0.1136	0.1132	0.1143	0.1127	0.1118	0.1125	0.1094	0.1094
RR	0.0627	0.0628	0.0617	0.0634	0.0755	—	—	—

图 6.29、表 6.9 及图 6.30、表 6.10 的预测结果表明，当邻点数远大于嵌入维数 m(局部线性预测)或待定系数个数 J(局部多项式预测)时，利用最小二乘法与

正则化方法的多变量局部线性预测法和多变量局部多项式预测法对上海证券市场综合股指的预测误差均较小，且正则化法的预测结果均要优于最小二乘法的预测结果。当邻点数接近或小于嵌入维数 m 或待定系数个数 J 时，利用最小二乘法与正则化方法的多变量局部线性预测法和多变量局部多项式预测法对上海证券市场综合股指的预测误差均有所增大，但利用最小二乘法的一般多变量局部线性预测法和多变量局部多项式预测法预测误差增幅最大且最终失效，而正则化方法中的 PCR 法仍有较小的预测误差。

上面利用了三种多变量非线性时间序列的预测法，即多变量局部多项式预测法、基于正则化的多变量局部线性预测法和基于正则化的多变量局部多项式预测法对上海证券市场的综合股指进行预测。预测结果表明这三种方法都是有效的，且表明多变量的局部多项式预测法预测效果要比单变量局部多项式预测法精确，基于正则化的多变量局部线性预测法和基于正则化的多变量局部多项式预测法的预测效果比一般的多变量局部线性预测法和多变量局部多项式预测法好，其抗噪性更强且应用范围更广泛。

参考文献

[1] Takens F. On the numerical determination of the dimension of an attractor. In: Rand D, Young L S editors. Dynamical Systems and Turbulence. Warwick, 1980, Lecture Notes in Mathematics, Springer-Verlag, 1981, 898: 366～381

[2] Mane R. On the dimension of the compact invariant sets of certain nonlinear maps. In: Rand D, Young L S editors. Dynamical Systems and Turbulence. Warwick, 1980, Lecture Notes in Mathematics, Springer-Verlag, 1981, 898: 230

[3] Sauer T, Yorke J A, Casdagli M. Embeddology. Journal of Statistical Physics, 1991, 65: 579～616

[4] Abarbanel H D I, Brown R, Sidorowich J J et al. The analysis of observed data in physical systems. Reviews of Modern Physics, 1993, 65(4): 1331～1392

[5] Fraser A M, Swinney H. Independent coordinates for strange attractors from mutual information. Physical Review A, 1986, 33(2): 1134～1140

[6] Fraser A M. Information and entroy in strange attractors. IEEE Trans. Inf. Theory, 1989, 35: 245～262

[7] Martinerie J M, Albano A M, Mees A I et al. Mutual information, strange attractors, and the optimal estimation of dimension. Physical Review A, 1992, 45(10): 7058～7064

[8] Rosenstein M T, Collins J J, DeLuca C J. Reconstruction expansion as a geometry-based framework for choosing proper delay times. Physica D, 1994, 73: 82～98

[9] Albano A M, Passamante A, Farrell M E. Using higher-order correlation to define an embedding window. Physica D, 1991, 54: 85～97

[10] Buzug T, Pfister G. Optimal delay time and embedding dimension for delay time coordinates by analysis of the global static and local dynamical behavior of strange attractors. Physical Review A, 1992, 45(10): 7073～7084

[11] Ding M Z, Grebogi C, Ott E et al. Estimating correlation dimension from a chaotic time series: when does plateau occur? Physica D, 1993, 69: 404～424

[12] Grassberger P, Procaccia I. Characterization of strange attractors. Physical Review Letters, 1983, 50(5): 346～349

[13] Kennel M B, Brown R, Abarbanel H D I. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Physical Review A, 1992, 45(6): 3403～3411

[14] Abarbanel H D I. Analysis of Observed Chaotic Data. New York: Springer-Verlag Inc., 1996

[15] Cao LY. Practical method for determining the minimum embedding dimension of a scalar

time series. Physica D, 1997, 110(5): 43~50

[16] Kim H S, Eykholt R, Salas J D. Nonlinear dynamics, delay times and embedding windows. Physica D, 1999, 127: 48~60

[17] Broomhead D S, King G P. Extracting qualitative dynamics from experimental data. Physica D, 1986, 20: 217~236

[18] Grassberger P, Procaccia I. Measuring the strangeness of strange attractors. Physica D, 1983, 9: 189~208

[19] Grassberger P, Procaccia I. Estimation of the Kolmogorov entropy from a chaotic signal. Physical Review A, 1983, 28(4): 2591~2593

[20] Wolf A, Swift J B, Swinney H L et al. Determining Lyapunov exponents from a time series. Physica D, 1985, 16: 285~317

[21] Theiler J. Estimating fractal dimension. J. Opt. Soc. Am. A, 1990, 7(6): 1055~1073

[22] Pesin Y B. On rigorous mathematical definitions of correlation dimension and generalized spectrum for dimension. Journal of Statistical Physics, 1993, 71(3/4): 529~547

[23] Schepers H E, Van Beek J H G M, Bassingthwaighte J B. Four method to estimation the fractal dimension from self-affine signals. IEEE Engineering in Medicine and Biology, 1992, 11(3): 57~64

[24] Rapp P E, Albano A M, Schmah T I. Filtered noised can mimic low-dimensional chaotic attractors. Physical Review E, 1993, 47(4): 2289~2297

[25] Osborne A R, Provenzale A. Finite correlation dimension for stochastic systems with power-law spectra. Physica D, 1989, 35: 357~381

[26] Lai Y C, Lerner D. Effective scaling regime for computing the correlation dimension from chaotic time series. Physica D, 1998, 115: 1~18

[27] Theiler J. Spurious dimension from correlation algorithms applied to limited time-series data. Physical Review A, 1986, 34(3): 2427~2432

[28] Theiler J. Statistical precision of dimension estimators. Physical Review A, 1990, 41(6): 3038~3051

[29] Judd K. An improved estimator of dimension and some comments on providing confidence intervals. Physica D, 1992, 56: 216~228

[30] Pisarenko D V, Pisarenko V F. Statistical estimation of the correlation dimension. Physics Letters A, 1995, 197: 31~39

[31] Lai Y C, Lerner D. Effective scaling regime for computing the correlation dimension from chaotic time series. Physica D, 1998, 115: 1~18

[32] Galka A, Maab T, Pfister G. Estimating the dimension of high-dimensional attractors: a comparison between two algorithms. Physica D, 1998, 121: 237~251

[33] Widman G, Lehnertz K, Jansen P et al. A fast general purpose algorithm for the computation of auto- and cross-correlation integrals from single channel data. Physica D, 1998, 121: 65~74

[34] Kitoh S, Kimura M, Mori T et al. A fundamental bias in calculating dimensions from finite data sets. Physica D, 2000, 141: 171～182

[35] Theiler J. Efficient algorithm for estimating the correlation dimension from a set of discreet point. Physical Review A, 1987, 36: 4456～4462

[36] Eckmann J P, Ruelle D. Fundamental limitation for estimating dimension and Lyapunov exponents in dynamical systems. Physica D, 1992, 56: 185～187

[37] Pawelzik K, Schuster H G. Generalized dimensions and entropies from a measured time series. Physical Review A, 1987, 35(1): 481～484

[38] Grassberger P. Finite sample corrections to entropy and dimension estimates. Physical Letters A, 1988, 128(6,7): 369～373

[39] Caputo J G, Atten P. Metric entropy: an experimental means for characterizing and quantifying chaos. Physical Review A, 1987, 35(3): 1311～1316

[40] Frank M, Blank H R, Heindl J et al. Improvement of K_2-entropy calculations by means of dimension scaled distances. Physica D, 1993, 65: 359～364

[41] Eckmann J P, Kamphorst S O, Ruelle D et al. Lyapunov exponents from time series. Physical Review A, 1986, 34(6): 4971～4979

[42] Brown R, Bryant P, Abarbanel H D I. Computing the Lyapunov spectrum of a dynamical system from an observed time series. Physical Review A, 1991, 43(6): 2787～2806

[43] 杨绍清,章新华,赵长安.一种最大李雅普诺夫指数估计的稳健算法.物理学报,2000,49(4):636～640

[44] Wu Z B. Remark on metric analysis of reconstructed dynamics from chaotic time series. Physica D, 1995, 85: 485～495

[45] Tanaka T, Aihara K, Taki M. Analysis of positive Lyapunov exponents from random time series. Physica D, 1998, 111: 42～50

[46] Oiwa N N, Ferrara N F. A fast algorithm for estimating Lyapunov exponents from time series. Physics Letters A, 1998, 246: 117～121

[47] Bunner M J, Hegger R. Estimation of Lyapunov spectra from space time data. Physics Letters A, 1999, 258: 25～30

[48] Ziehmann C, Smith L A, Kurths J. The bootstrap and Lyapunov exponents in deterministic chaos. Physica D, 1999, 126: 49～59

[49] 马军海,陈予恕,刘曾荣. 动力系统实测数据的 Lyapunov 指数的矩阵算法. 应用数学与力学,1999,20(9):919～926

[50] 马军海,陈予恕,季进臣. 三种动力系统 Lyapunov 指数的比较. 天津大学学报,1999,32(2):190～195

[51] Diks C, Van Houwelingen J C, Takens F et al. Reversibility as a criterion for discriminating time series. Physics Letters A, 1995, 201: 221～228

[52] Blanco S, Garcia H, Quian R et al. Stationary of the EEG series. IEEE Engineering in Medicine and Biology, 1995, 14(4): 395～399

[53] Jenkins G M, Watls D. Spectral Analysis And Its Applications. San Francisco: Holden-Day, 1968

[54] Priestley M B. Non-Linear And Non-Stationary Time Series Analysis. London: Academic Press, 1988

[55] Isliker H, Kurths J. A test for stationarity: finding parts in time series apt for correlation dimension estimate. Int. J. Bifurcation Chaos, 1993, 3: 1573～1579

[56] Provenzale A, Smith L A, Vio R et al. Distinguishing between low-dimensional dynamics and randomness in measured time series. Physica D, 1992, 58: 31～49

[57] Kennel M B. Statistical test for dynamical nonstationarity in observed time-series data. Physical Review E, 1997, 56(1): 316～321

[58] Eckmann J P, Kamphost S O, Ruelle D. Recurrence plots of dynamical systems. Europhysics Letters, 1987, 4(9): 973～977

[59] Casdagli M C. Recurrence plots revisited. Physica D, 1997, 108: 12～44

[60] Schreiber T. Detecting and analyzing nonstationarity in a time series using nonlinear cross predictions. Physical Review Letters, 1997, 78(5): 843～846

[61] Kantz H. Quantifying the closeness of fractal measures. Physical Review E, 1994, 49 (6): 5091～5097

[62] Theiler J, Galdrikan B, Longtin A et al. Using surrogate data to detect nonlinearity in time series, in: Nonlinear Modeling and Forecasting, SFI Studies in the Sciences of Complexity, Proc. Vol. XII, eds. Casdaglis and S. Eubank (Addision-Wesley, Reading, MA), 1992, 163～188

[63] Theiler J, Eubank S, Longtin A et al. Testing for nonlinearity in time series: the method of surrogate data. Physica D, 1992, 58: 77～94

[64] 马军海,混沌时序非线性动力系统重构．天津大学博士学位论文,1997

[65] Stam C J, Pijn J P M, Pritchard W S. Reliable detection of nonlinearity in experimental time series with strong periodic components. Physica D, 1998, 112: 361～380

[66] Schreiber T, Schmitz A. Improved surrogate data for nonlinearity tests. Physical Review Letters, 1996, 77(4): 635～638

[67] Kugiumtzis D. Test your surrogate data before you test for nonlinearity. Physical Review E, 1999, 60(3): 2808～2816

[68] Kugiumtzis D. Surrogate data test for nonlinearity including nonmontonic transforms. Physical Review E, 2000, 62(1): R25～R28

[69] 刘耀宗,温熙森,胡茑庆．非最小相位线性非高斯序列的替代数据检验．物理学报,2001, 50(4):633～637

[70] 刘耀宗,温熙森,胡茑庆．线性非高斯序列的替代数据检验新方法．物理学报,2001,50 (7):1241～1246

[71] Scjreiber T. Constrained randomization of time series data, Physical Review Letters, 1998, 80(10): 2105～2108

[72] Schreiber T. Schmitz A. Discrimination power of measures for nonlinear in a time series. Physical Review E, 1997, 55(5): 5543~5547

[73] Schreiber T, Schmitz A. Surrogate time series. Physica D, 2000, 142: 346~382

[74] Kaplan D, Glass L. Understanding Nonlinear Dynamics. New York: Springer-Verlag Inc., 1997

[75] Eckmann J P, Kamphorst S O, Ruelle D. Recurrence plots of dynamical systems. Europhysics Letters, 1987, 4(9): 973~977

[76] Casdagli M C, Recurrence plots revisited. Physica D, 1997, 108: 12~44

[77] Choi J M, Bae B H, Kim S Y. Divergence in perpendicular recurrence plot: quantification of dynamical divergence from short chaotic time series. Physics Letters A, 1999, 263: 299~306

[78] Holyst J A, Zebrowska M. Observations of deterministic chaos in financial time series by recurrence plots, can one control chaotic economy? The European Physical Journal B, 2001, 20: 531~535

[79] Marwan N, Kurths J. Line structures in recurrence plots. Physics Letters A, 2005, 336 (4-5): 349~357

[80] Marwan N, Kurths J. Cross recurrence plots and their applications. In: Benton C V Ed. Mathematical Physics Research Cutting Edge, Hauppauge. Nova Science Publishers, 2004, 101~139

[81] Zbilut J P, Giuliani A, Colosimo A, Mitchell J C et al. Charge and hydrophobicity patterning along the sequence predicts the folding mechanism and aggregation of proteins: A computational approach. Journal of Proteome Research, 2004, 3: 1243~1253

[82] Zbilut J P, Mitchell J C, Giuliani A et al. Singular hydrophobicity patterns and net charge: A mesoscopic principle for protein aggregation/folding. Physica A, 2004, 343: 348~358

[83] Schreiber T. Determination of the noise level of a chaotic time series. Physical Review E, 1993, 48(1): R13~R16

[84] Schreiber T. Influence of Gaussian noise on the correlation exponent. Physical Review E, 1997, 56(1): 274~277

[85] Urbanowicz K, Holyst J A. Noise-level estimation of time series using coarse-grained entropy. Physical Review E, 2003, 67: 046218(1)~046218(7)

[86] Yu D, Small M, Harrison R G et al. Efficient implementation of the Gaussian kernel algorithm in estimating invariants an noise level from noisy time series data. Physical Review E, 2000, 61(4): 3750~3756

[87] Schreiber T. Extremely simple nonlinear noise-reduction method. Physical Review E, 1993, 47(4): 2401~2404

[88] Kostelich E J, Schreiber T. Noise reduction in chaotic time series data: a survey of common methods. Physical Review E, 1993, 48(3): 1752~1763

[89] Schreiber T, Kantz H. Noise in chaotic data: diagnosis and treatment. Chaos, 1995, 5. 133～142

[90] Schreiber T. Grassberger P. A simple noise reduction method for real data. Physics Letters A, 1991, 160: 411～418

[91] Kugiumtzis D. Correction of the correlation dimension for noisy time series. International Journal of Bifurcation and Chaos, 1997, 7: 1283～1294

[92] Kantz H, Schreiber T. Nonlinear Time Series Analysis. Cambridge University Press, 1997

[93] Palus M, Albrecht V, Dvorak I. Information-theoretic test for nonlinearity in time series. Physics Letters A, 1993. 175: 203～209

[94] Palus M. Testing for nonlinearity using redundancies: quantitative and qualitative aspects. Physica D, 1995, 80: 186～205

[95] Palus M. Detecting nonlinearity in multivariate time series. Physics Letters A, 1996, 213: 138～147

[96] 王海燕,盛昭瀚,李文. 多维信号非线性特征的检验. 信号处理,2002,18(1):20～23

[97] 王海燕,汤龙坤. 一种定量检验多维信号非线性的方法. 信号处理,2003,19(5):407～410

[98] Wang H Y, Tang L K. Testing for nonlinear in Shanghai stock market. International Journal of Modern Physics B, 2004, 18(17-19): 2720～2724

[99] Wang H Y, Tang L K. Detecting nonlinearity in stock trading volume time series. Proceedings of 2004 International Conference on Management Science & Engineering. edited by Lan Hua. Harbin Institute of Technology Press, 2004, 2128～2132

[100] Prichard D, Theiler J. Generalized redungancies for time series analysis. Physica D, 1995, 84: 476～493

[101] Press W H, Flannery B P, Teukolsky S A et al. Numerical Recipes: the Art of Scientific Computing. Cambridge, M. A.: Cambridge University Press, 1986

[102] Prichard D, Theiler J. Generating surrogate data for time series with several simultaneously measured variables. Physical Review Letters, 1994, 73(7): 951～954

[103] 盛昭瀚,姚洪兴,王海燕,陈国华. 混沌动力系统的重构预测与控制. 北京:中国经济出版社,2003

[104] Pomp B. Measuring statistical dependences in a time series. Journal of Statistical Physics, 1993, 73(3/4): 587～610

[105] Pomp B, Blidh P, Hoyer D et al. Using mutual information to measure coupling in the cardiorespiratory system. IEEE Engineering in Medicine and Biology, 1998, 17(6): 32～39

[106] 王海燕,李文,陈文彦. 时间序列间统计依赖性测量的几种方法及比较. 南京理工大学学报,2002,26(3):325～329

[107] 王海燕,盛昭瀚,陈文彦. 混沌系统耦合强度的一种测量方法. 非线性动力学学报,2002,9(1/2):19～25

[108] Eckmann J P , Ruelle D . Ergodic theory of strange attractors. Review Modern Physics, 1985, 57: 617～656

[109] Destexhe A, Sepulcher J A, Babloyantz A. A comparative study of the experiment quantification of deterministic chaos. Physics Letters A, 1988, 132: 101～106

[110] Rombowts S A R B, Keunen R W M, Stom C J. Investigation of nonlinear structure in multichannel EEG. Physics Letters A, 1995, 202: 352～358

[111] Dvorak I. Takens versus multichannel reconstruction in EEG correlation exponent estimates. Physics Letters A, 1990, 151(5): 225～233

[112] Cao LY, Mees A, Judd K. Dynamics from multivariate time series. Physica D, 1998, 121: 75～88

[113] Boccaletti S, Valladares D L. Reconstructing embedding spaces of coupled dynamical systems from multivariate data. Physical Review E, 2002, 65:1～4

[114] 王海燕,盛昭瀚,张进.多变量时间序列复杂系统的相空间重构.东南大学学报(自然科学版),2003,33(1):115～118

[115] 卢山,王海燕.多变量时间序列最大李雅普诺夫指数的计算.物理学报,2006,55(2):572～576

[116] Rosenblum M G, Pikovsky A S. Phase synchronization of chaotic oscillators. Physical Review Letters, 1996, 76: 1804～1807

[117] 赵贵兵,石炎福,段文锋,余华瑞.从混沌时间序列同时计算关联维和 Komogorov 熵.计算物理,1999,16(3):309～315

[118] Rosenstein M T,Collins J J,De Luca C J. A practical method for calculating largest Lyapunov exponents from small data sets. Physica D, 1993, 65: 117～134

[119] 杨绍清,贾传荧.两种实用的相空间重构方法.物理学报,2002,51(11):2452～2456

[120] Farmer J D, Sidorwich J J. Predicting Chaotic Time Series. Physical Review Letters, 1987, 59(8): 845～848

[121] 吴耿锋,周佩玲,储阅春等.基于相空间重构的预测方法及其在天气预报中的应用.自然杂志,1999,21(2):107～110

[122] 向小东,郭耀煌.基于混沌吸引子的时间序列预测方法及其应用.西南交通大学学报,2001,36(5):472～475

[123] Jayawardena A W, Li W K, Xu P. Neighborhood Selection for local modelling and prediction of hydrological time series. Journal of Hydrology, 2002, 258: 40～57

[124] Jianming Y. On measuring and correcting the effects of data mining and model selection. J. Am Stat. Assoc., 1998, 93(441): 120～131

[125] Farmer J D , Sidorowich J J. Exploiting chaos to predict the future and reduce noise. In: Lee Y C Ed. Evolution, Learning and Cognition. Singapore: World Scientific, 1988, 277～330

[126] Itoh K. A method for predicting chaotic time-series with outliers. Electronics and Communications in Japan, 1995, 78(5): 44～53

[127] Kasim K, Levent S, Orhan S. Nonlinear time series prediction of O3 Concentration in Istanbul. Atmospheric Environment, 2000, 34(8): 1267～1271

[128] Fang F, Wang H Y. Local polynomial prediction method of multivariate chaotic time series and its application. Journal of Southeast University (English Edition), 2005, 21 (2), 229～232

[129] 吕金虎,陆君安,陈士华. 混沌时间序列分析及其应用. 武汉:武汉大学出版社,2002

[130] 王永忠,曾昭磐. 混沌时间序列的局域线性回归预测方法. 厦门大学学报,1999,38(4):636～640

[131] 白晓林,胡兴锐,徐雄. 混沌时间序列局域线性预测方法. 上海交通大学学报,1999,33(1):19～21

[132] Linsay P S. An efficient method of forecasting chaotic time series using linear interpolation. Physics Letters A, 1991, 153(6,7): 353～356

[133] Navone H D, Ceccatto H A. Forecasting Chaos from small data sets: a comparison of different nonlinear algorithms. J Phys, A: Math Gen, 1995, 28(12): 3381～3388

[134] Gencay R. Nonlinear prediction of noisy times series with feedforward networks. Physics Letters A, 1994, 187: 397～403

[135] de Oliveira K A, Vannucci A, da Silva E C. Using artificial neural networks to forecast chaotic time series. Physica A, 2000, 284: 393～404

[136] Lisi F, Schiavo R A. A comparison between neural networks and chaotic models for exchange rate prediction. Computational Statistics & Data Analysis, 1999, 30: 87～102

[137] 顾玉巧,陈天仑,黄五群. 基于混沌算法神经网络的恒生指数期货时间序列预测. 南开大学学报,2000,33(1):119～123

[138] 杨一文,刘贵忠,张宗平. 基于嵌入理论和神经网络技术的混沌数据预测及其在股票市场中的应用. 系统工程理论与实践,2000,21(6):52～58

[139] 李克平,陈天仑. 混沌算法神经网络与含噪声时间序列的预测. 南开大学学报,2001,34(3):28～31

[140] 顾炜,瞿东辉,张立明. 对复杂混沌时间序列快速预测的前馈神经网络. 复旦学报,1995,34(3):262～268

[141] Gencay R, Tung L. Nonlinear modeling and prediction with feedforward and recurrent network. Physica D, 1997, 108: 119～134

[142] Cao L, Hong Y, Fang H, He G, Predicting chaotic time series with wavelet networks. Physica D, 1995, 85(1): 225～238

[143] 陈哲,冯天瑾,张海燕. 基于小波神经网络的混沌时间序列分析与相空间重构. 计算机研究与发展,2001,38(5):592～596

[144] Maguire L P, Roche B, Mcginnity T M et al. Prediction a chaotic time series using a fuzzy neural network. Information Sciences, 1998, 112: 125～136

[145] 郭刚,史忠科,戴冠中. 基于混沌理论进行股票市场的多步预测. 信息与控制,2000,29(2):177～181

[146] 郭刚,史忠科,戴冠中.依据混沌理论进行非线性系统建模变量个数的最优选取.控制与决策,2000,15(2):233～238

[147] 徐秉铮,张百灵,韦岗.神经网络理论与应用.广州:华南理工大学出版社,1994

[148] 王上飞,周佩玲,吴耿峰等.径向基神经网络在股市预测中的应用.预测,1998,(6):44～46

[149] 向小东,郭耀煌.基于径向基函数网络的改进算法的股市数据预测.预测,2002,21(4):66～68

[150] 王明进,程乾生.Kohonen 自组织神经网络在混沌时间序列预测中的应用.系统工程理论与实践,1997,(7):12～17

[151] 王明进,程乾生.基于径向基函数的非线性预测模型.管理科学学报,1999,2(4):28～33

[152] 程乾生,武连生,王守章.基于属性聚类网络和径向基函数的融合预测.科学通报,2000,45(11):1211～1216

[153] Casdagli M. Nonlinear prediction chaotic time series. Physica D, 1989, 35: 335～356

[154] Broomhead D S, Lowe D. Multivariable functional interpolation and adaptive networks. Complex Systems, 1988, 2: 321～355

[155] Wang H Y, Zhu M. A prediction comparison between univariate and multivariate chaotic time series. Journal of Southeast University (English Edition), 2003, 19(4): 414～417

[156] Scharf L L. Statistical Signal Processing: Detection, Estiamation and Time Series Analysis. Addison-Wesley. Reading. MA, 1991

[157] Kugiumtzis D, Lingjerde O C, Christophersen N. Regularized local linear prediction of chaotic time series. Physica D, 1998, 112(3-4): 344～360

[158] Massy W F. Principal components regression in exploratory statistical research. Journal of the American Statistical Association, 1965, 60: 234～246

[159] Lukas M A. Comparisons of parameter choice methods for regularization with discrete noisy data. Inverse Problems, 1998, 14: 161～184

[160] 孙文爽,陈兰祥.多元统计分析.北京:高等教育出版社,1994

[161] Hoerl A E, Kennard R E. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 1970, 12(1): 55～109

[162] Wold S, Albano C. Analysis of two partial least squares algorithms for multivariate calibration. Chemometrics and Intelligent Laboratory Systems. 1983, 2: 187～197

[163] 李爱国,覃征.自适应局部线性化法预测混沌时间序列.系统工程理论与实践,2004,6(6):67～71

[164] 何晓群等.应用回归分析.北京:中国人民大学出版社,2001

[165] 埃德加·E·彼得斯.资本市场的混沌与秩序.王小东译.北京:经济科学出版社,1999

[166] 埃德加·E·彼得斯.分形市场分析——将混沌理论应用到投资与经济理论上.储海林,殷勤译.北京:经济科学出版社,2002

[167] Chen P. A random walk or color chaos on the stock market? Time-frequency analysis of S&P indexes. Studies in Nonlinear Dynamics & Econometrics,1996,1(2): 87～103

[168] Chen P. Study of chaotic dynamical systems via time-frequency analysis. Proceedings of IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis. Piscataway, New Jersey, IEEE, 1994, 357～360

[169] 卢山,王海燕.证券市场的非线性和确定性检验.系统工程理论方法应用,2005,14(3):235～238

[170] 赵巍,王海燕,钱志强.沪深股票市场非线性特征和星期效应的R/S分析.湖北大学学报(自然科学版),2003,25(专辑):79～83

[171] 张维,黄兴.沪深股市的R/S实证分析.系统工程,2001,19(1):1～5

[172] 徐龙炳,陆蓉.R/S分析探索中国股票市场的非线性.预测,1999,(2):59～62

[173] 刘孝贤,郭举修.上证综合指数的非线性动力学特点.山东工业大学学报,2001,31(5):416～423

[174] 段虎,沈菲.上证指数中的混沌现象研究.数量经济技术经济研究,2002,(2):91～94

[175] 朱梅,王海燕.中国股票市场的非线性确定性预测.安徽工程科技学院学报,2004,19(2):10～13

[176] 卢山,王海燕.金融时间序列局部预测算法的优化研究.2005中国控制与决策学术年会论文集,2005,1:1808～1811

[177] Lu S, Charles N P. Optimized local constant prediction with application on financial time series. Proceedings of 2004 international conference on management science & engineering, 2004,1: 2160～2165

[178] 王海燕,盛昭瀚.混沌时间序列相空间重构参数的选取方法.东南大学学报(自然科学版),2000,30(5):113～117